Modelling impulsive noise in indoor powerline communication systems

Oktay Karakuş1,4 · Ercan E. Kuruoğlu2,3 · Mustafa A. Altınkaya4

Received: 29 July 2019 / Revised: 29 January 2020 / Accepted: 2 May 2020 / Published online: 4 June 2020
© The Author(s) 2020

Abstract

Powerline communication (PLC) is an emerging technology that has an important role in smart grid systems. Due to making use of existing transmission lines for communication purposes, PLC systems are subject to various noise effects. Among those, the most challenging one is the impulsive noise compared to the background and narrowband noise. In this paper, we present a comparative study on modelling the impulsive noise amplitude in indoor PLC systems by utilising several impulsive distributions. In particular, as candidate distributions, we use the symmetric $\alpha$-Stable ($S\alpha S$), generalised Gaussian, Bernoulli Gaussian and Student’s $t$ distribution families as well as the Middleton Class A distribution, which dominates the literature as the impulsive noise model for PLC systems. Real indoor PLC system noise measurements are investigated for the simulation studies, which show that the $S\alpha S$ distribution achieves the best modelling success when compared to the other families in terms of the statistical error criteria, especially for the tail characteristics of the measured data sets.

1 Introduction

Smart grid technology, which is regarded as the next generation electric power infrastructure, is being developed day by day in order to provide a more efficient and safe power system. The ability to control this complex system automatically (or remotely) also makes it necessary to use advanced communication technologies [1]. Powerline communication (PLC) systems use power lines to carry telecommunication data. In PLC systems, a communication speed can be achieved up to 2Gbps with a good quality of service. PLC systems have important potential in applications, namely remote metering, distribution automation and internet access through home networking [2]. Moreover, PLC can also provide a physical environment for closed multimedia data traffic without additional cables. Despite all these competencies, the most challenging problem of PLC systems is the transmission of data in an environment designed for electricity distribution. For this reason, in PLC systems, a more complicated noise environment arises when compared to conventional communication systems.

A PLC system has various types of noise arising from electrical devices connected to power line as well as external effects via electromagnetic radiation. These noise sequences are generally non-Gaussian and are classified into three groups, namely i) impulsive noise, ii) narrowband noise and iii) background noise [3]. Impulsive noise in PLC systems is the most common cause of communication errors compared to other types of noise due to its high amplitudes which can exceed 50 dBs [4,5].

In the literature, Middleton Class A distribution [6] is the most common choice in modelling impulsive noise in PLC systems [4,7]. This statistical model is a mixture of a large number of Gaussian random variables with different vari-
ances. Middleton Class A distribution has had a widespread usage in PLC systems due to the simplicity of its probability density function (pdf). $\alpha$-Stable distributions have emerged as an important alternative to Middleton model in modelling impulse noise in recent years, mainly due to the fact that Middleton Class A and Bernoulli-Gaussian (BG) distributions cannot model noise tail characteristics well. $\alpha$-Stable distributions have been proposed to model PLC impulsive noise in [8]. Furthermore, in [9] a mixture model of BG and distributions have been proposed to model PLC impulsive noise. Middleton Class A and Bernoulli-Gaussian (BG) distributions have been used for modelling the narrowband PLC systems. In [10], Middleton Class A and $\alpha$-Stable distributions are used to model PLC noises in China and Italy, and it has been stated that PLC systems including high impulsive character can be modelled better with $\alpha$-Stable distributions.

Lastly, in [11], the suitability of Middleton Class A distribution has been evaluated under various conditions, and it has been shown that the Middleton Class A distribution has quite limited usage in PLC systems, due to failure in modelling the tail characteristics of the impulsive noise.

For the PLC system applications, there remains the question what the most suitable distribution for impulse noise modelling is, as none of the above-mentioned studies have provided a comparison that covers all frequently used impulsive distributions in the literature.

In this paper, we contribute to the literature with:

1. a comparison study among various impulsive distribution families instead of focusing on a single distribution. Candidate distributions are namely the Middleton Class A, BG, $\alpha$S, Student’s $t$ and generalised Gaussian (GG).
2. an investigation on four different indoor PLC noise measurements of various indoor locations of namely a house, a room at university and two laboratories [12].
3. a statistical significance test among said distribution families on modelling the real noise measurements in terms of statistical error measures of Kolmogorov–Smirnov (KS) statistics, Kullback–Leibler (KL) divergence and Hellinger distance (HD).
4. an experimental verification to the studies [8,13] which propose utilising the $\alpha$S distribution for modelling PLC impulsive noise under the lights of the simulation results.
5. an alternative distribution family, namely the Student’s $t$, which would be applicable in PLC impulsive noise modelling with its considerable modelling performance and simpler mathematical expression compared to the $\alpha$S distribution.

In this paper, we only focus on the amplitude of noise measurements. Thus, please note that the other important characteristics of the impulsive noise such as inter-arrival times, widths and frequencies are out of the scope of this study and left as future work.

The rest of the paper is organised as follows: candidate probability distributions, parameter estimation method and error measures are presented in Sect. 2. The details of the data sets and the experimental analysis results are discussed in Sect. 3. Section 4 provides the conclusions of the paper.

## 2 Methodology

### 2.1 Probability distributions

As mentioned in Sect. 1, we investigate five different distribution families in modelling the impulsive noise of PLC systems in this study. The corresponding five candidate distribution families are Middleton Class A, BG, $\alpha$S, Student’s $t$ and GG. A brief information about these distributions is given in the sequel.

#### 2.1.1 Middleton Class A model

Middleton Class A distribution can be defined as a Gaussian mixture distribution, weights of which are Poisson random variables. The pdf of the Middleton Class A model is expressed as

$$f(x) = \sum_{m=0}^{\infty} P_m \mathcal{N}(x|0, \sigma_m^2)$$

(1)

where $\sigma_m^2$ is the variance and $P_m$ refers to the weight of the $m$th Gaussian. Both of these parameters are given as

$$\sigma_m^2 = \sigma^2 \left(\frac{m + \gamma}{A + \gamma}\right) \quad \text{and} \quad P_m = \frac{e^{-A} A^m}{m!},$$

(2)

where $\sigma^2$ represents the variance of the background noise (generally assumed to be 1), whilst $A$ refers to the shape parameter and $\gamma$ is the scale parameter.

In previous studies [11,14], it has been shown that the Middleton Class A distribution can be simplified into a summation up to a finite scalar $M$, instead of performing the infinite summation in (1). In this study, we assume that $M$ is 10, and the weight vector, $P_m$ in (1), is replaced with

$$P_m' = \frac{P_m}{\sum_{i=0}^{10} P_m},$$

For details, please see [6].

#### 2.1.2 Bernoulli-Gaussian distribution

The BG distribution is another important statistical model for PLC noise modelling. It is a two-component Gaussian mixture model and also a special case of Middleton Class A pdf with first two components (i.e. for $m = 0, 1$ in (1)) [11].
The BG model has various usage areas, especially in communication applications such as PLC [15,16], multicarrier QAM [17] and orthogonal frequency division multiplexing (OFDM) systems [18].

The pdf expression of the BG distribution is given as [16]

\[
\begin{align*}
  f(x) &= \frac{1 - p}{\sqrt{2\pi\sigma_B^2}} \exp \left(\frac{-x^2}{2\sigma_B^2}\right) \\
  &\quad + \frac{p}{\sqrt{2\pi(\sigma_B^2 + \sigma_I^2)}} \exp \left(\frac{-x^2}{2(\sigma_B^2 + \sigma_I^2)}\right)
\end{align*}
\]

where \( p \in [0, 1] \) refers to the impulsive probability, \( \sigma_B^2 \) is the standard deviation of the background noise amplitude and \( \sigma_I^2 \) represents the standard deviation of the impulsive noise amplitude.

2.1.3 Symmetric \( \alpha \)-stable distribution

\( \alpha \)-Stable distribution is a commonly used impulsive distribution with various applications such as near optimal receiver design [19] and diversity combining schemes for a single-user communication [20]. (Please see [21] and references therein for detailed applications.)

The \( \alpha \)S\( \alpha \)S distribution has no closed form pdf expression except for the special members the Cauchy (\( \alpha = 1 \)) and the Gaussian (\( \alpha = 2 \)). However, its characteristic function can be expressed explicitly as:

\[
\phi(t) = \exp(j\delta t - \gamma |t|^{\alpha})
\]

where \( 0 < \alpha \leq 2 \) refers to the shape parameter (or namely the characteristic exponent), which controls the impulsiveness of the distribution. \( -\infty < \delta < \infty \) is the location parameter, whilst \( \gamma > 0 \) is the scale parameter (or namely the dispersion) and can be expressed as a measure of the spread of the distribution around \( \delta \).

2.1.4 Student’s \( t \) distribution

Student’s \( t \) distribution is an alternative to Gaussian distribution, especially for small populations where the validity of central limit theorem is questionable. Student’s \( t \) distribution has various application areas, to name a few: interference modelling in cognitive radio network [22] and analysis of enhanced noise after zero-forcing frequency domain equalisation [23].

The univariate symmetric Student’s \( t \) distribution is defined with three parameters: (i) the shape parameter \( \alpha > 0 \) (namely the number of degrees of freedom), (ii) the location parameter \( -\infty < \delta < \infty \) and (iii) the scale parameter \( \gamma > 0 \). As in the \( \alpha \)S\( \alpha \)S case, the Cauchy and Gauss distributions are the special members of the Student’s \( t \) distribution, which are obtained for shape parameter values of \( \alpha = 1 \) and \( \alpha = \infty \), respectively.

The pdf expression for the Student’s \( t \) distribution is given as

\[
\begin{align*}
  f(x) &= \frac{\Gamma\left(\frac{\alpha + 1}{2}\right)}{\Gamma(\alpha/2)\sqrt{\pi\alpha}} \left(1 + \frac{1}{\alpha} \left|\frac{x - \delta}{\gamma}\right|^2\right)^{-\left((\alpha+1)/2\right)},
\end{align*}
\]

where \( \Gamma(\cdot) \) represents the gamma function.

2.1.5 Generalised Gaussian distribution

GG distribution has generally been used in applications such as performance analysis studies in multihop wireless networks [24] and M-PAM/M-QAM systems [25]. The pdf expression for the univariate GG distribution is given as

\[
\begin{align*}
  f(x) &= \frac{\alpha}{2\gamma \Gamma(1/\alpha)} \exp\left(-\left|\frac{x - \delta}{\gamma}\right|^\alpha\right)
\end{align*}
\]

where \( \alpha > 0 \) is the shape parameter, \( -\infty < \delta < \infty \) refers to the location parameter and the \( \gamma > 0 \) is the scale parameter. The Laplace, Gaussian and uniform distributions are special members of the GG family for \( \alpha \) values of 1, 2 and \( \infty \), respectively.

2.2 Parameter estimation method

In this study, in order to estimate parameters for all the candidate distributions, we use a minimax methodology as suggested in [11]. In particular, we utilise the cumulative density function (CDF) in the loss function of the minimax problem, inasmuch as the CDF contains all the information for a distribution of a random variable.

Given data \( x \), and the empirical CDF \( F(x) \) is readily obtained. The parameter vector \( \Theta \) of any distribution which gives the best fit to the given data \( x \) is then subsequently estimated via [11]:

\[
\hat{\Theta}_{\text{CDF}} = \arg\min_{\Theta} \max_x \left| F(x) - \hat{F}(x; \Theta) \right|
\]

where \( \hat{\Theta}_{\text{CDF}} \) refers to the parameter vector estimate which minimises the maximum distance between the empirical and the estimated CDFs. Here, please note that location parameters for Student’s \( t \) and GG distributions are assumed to be 0, since all other statistical models are symmetric around 0.

In order to solve (7), we employ an interior-point optimisation method which approaches the optimum solution from the interior of the feasible set. Particularly, it is a constrained optimisation method and the constraints are defined
for each of the shapes and the scale parameters, the upper and lower bounds of which are given in Sect. 2 for each distribution. The maximum number of iterations is selected as 100 which is experimentally set. We repeat solving the minimax problem in (7) 50 times for randomly selected initial points \((\Theta_0 = [\alpha_0, \gamma_0])\).

### 2.3 Statistical error measures

In order to demonstrate a statistical significance test among all distribution families for all the utilised data sets, we use three different statistical measures, which are the KS statistics, the KL divergence and the HD.

The two-sample KS statistics is a nonparametric test which can be used to test equality of continuous distributions by comparing two samples from two reference CDFs \(P(\cdot)\) and \(G(\cdot)\) and is expressed as [26]

\[
D_{\text{KS}} = \max \{ |P(x) - G(x)| \} . \tag{8}
\]

KL divergence provides a non-symmetric measure (not a distance) of the difference between information contents of two probability distributions. KL divergence is also known as the relative entropy and is calculated as [27]

\[
D_{\text{KL}} = \sum_{i=1}^{k} p(x_i) \log \left( \frac{p(x_i)}{g(x_i)} \right) . \tag{9}
\]

HD is a symmetric distance metric between two pdfs which measures how two distributions overlap. On the other hand, HD is strongly related to Bhattacharyya distance and also probabilistic analogous of Euclidean distance. HD is calculated as [28]

\[
D_{\text{HD}} = \frac{1}{\sqrt{2}} \| \sqrt{p(x)} - \sqrt{g(x)} \|_2 . \tag{10}
\]

When evaluating the results for all these statistical error measures, note that the smaller the statistical test value is, the higher the estimation performance is.

### 3 Experimental analysis

#### 3.1 Data sets

In this study, four different indoor PLC noise measurements are investigated, all of which have been measured during a project (PTDC/EEA-TEL/67979/2006) conducted by INESC/IST, Portugal [12]. There are four different measurement locations, which are namely a house, a university room and two laboratories (Lab1 and Lab2). Particularly, House, Room (a university room) and Lab2 measurements have been taken with the same type of amplifier, whereas Lab1 measurement has been taken with another amplifier with lower gain. Measurements were taken on December 2017. The sampling rate was 200 Msamples/s and measurements lasted for 5 ms resulting 1 million samples for each. For measurement set-up and further details of the data sets, please see [29].

Table 1: Estimated parameters for each family and data set

|       | House | Lab1 | Room | Lab2 |
|-------|-------|------|------|------|
| \(a\) | 1.1743| 1.2239| 1.2285| 1.1465|
| \(\hat{\gamma}\) | 2.5848| 2.6176| 2.6194| 2.5866|
| Student’s \(t\) | | | | |
| \(\hat{a}\) | 1.3765| 1.5239| 1.5239| 1.3066|
| \(\hat{\gamma}\) | 2.8551| 2.9632| 2.9672| 2.8164|
| GG | | | | |
| \(\hat{a}\) | 0.7242| 0.7737| 0.7855| 0.7005|
| \(\hat{\gamma}\) | 2.5960| 2.7740| 2.9739| 2.4882|
| Middleton | | | | |
| \(\hat{A}\) | 0.0201| 0.0251| 0.0229| 0.0182|
| \(\hat{\gamma}\) | 0.0047| 0.0210| 0.0056| 0.0027|
| BG | | | | |
| \(p\) | 0.2294| 0.2710| 0.8129| 0.4012|
| \(\sigma_B\) | 2.8937| 2.7728| 0.8501| 4.6115|
| \(\sigma_I\) | 19.6419| 15.0280| 6.4391| 97.7838|

Table 2: Computed statistical error measures for all the data sets

|       | House | Lab1 | Room | Lab2 |
|-------|-------|------|------|------|
| KS   | 0.0031| 0.0035| 0.0062| 0.0093| 0.0025|
| KL   | \textbf{0.0020}\ | 0.0027| 0.0105| 0.0960| 0.0059|
| HD   | \textbf{0.0339}\ | 0.0384| 0.0755| 0.1128| 0.0451|
| Middleton | | | | | |
| KS   | \textbf{0.0026}\ | 0.0032| 0.0060| 0.0089| 0.0035|
| KL   | \textbf{0.0018}\ | 0.0028| 0.0275| 0.1015| 0.0068|
| HD   | \textbf{0.0310}\ | 0.0376| 0.0745| 0.1028| 0.0539|
| Room | | | | | |
| KS   | \textbf{0.0025}\ | 0.0031| 0.0057| 0.0083| 0.0144|
| KL   | \textbf{0.0018}\ | 0.0028| 0.0283| 0.0993| 0.0209|
| HD   | \textbf{0.0308}\ | 0.0373| 0.0739| 0.1042| 0.0994|
| Lab2 | | | | | |
| KS   | \textbf{0.0034}\ | 0.0037| 0.0064| 0.0098| 0.0529|
| KL   | \textbf{0.0024}\ | 0.0029| 0.0212| 0.1022| 0.0298|
| HD   | \textbf{0.0369}\ | 0.0405| 0.0797| 0.1180| 0.1345|

Values with bold face refer to the best statistical measure values for each data sets and statistical measures.
3.2 Simulation results

In order to study the performance of the candidate distributions for each of the four PLC noise data sets, the required parameters were estimated using the method given in Sect. 2.2. The fitted distributions were calculated using the estimated parameters via their pdf expressions given in the above sections. PLC noise measurements are usually symmetrical around zero, and since all of the other candidate models are symmetrical distributions, for a fair comparison symmetrical sub-class of $\alpha$-Stable distribution family was chosen as a candidate distribution family ($\beta = \delta = 0$).

In Table 1, estimated parameters are shown for all distributions and data sets. Examining the estimated shape parameter values, e.g. for the SoS in Table 1, we can state that Lab2 and House measurements follow more impulsive characteristics compared to other two measurements. Additionally, Lab1 and Room data sets have the highest shape parameter estimates among the data sets and this corresponds to the least impulsiveness.

In Table 2, the performance comparison results are shown in terms of KS, KL, and HD statistical tests. The most important result that can be clearly seen from the table is the remarkable success of the SoS for all four data sets. Particularly, it has been observed that in all cases, SoS distribution has provided a better model in terms of visual comparison as well as KL, KS and HD except the single case of KS test for House data set. Even for this case, the visual comparisons of pdfs show the superior fit of SoS. These results are due to characteristics of the KS test which gives equal importance to all amplitude values and gives the maximum deviation between the two CDFs. A %1 deviation between the main peaks of the distributions might contribute the same amount with %50 deviation in the tails. KL distance on the other hand...
provides an information theoretic measure which scales the contribution to error according to the probabilities.

Furthermore, for all data sets Student’s $t$ distribution achieves the second best performance after $S\alpha S$. BG, GG and Middleton models fall short of modelling PLC noise measurements, especially for tail modelling, and this result also provides a comparative evidence to the results presented in [11] for Middleton Class A distributions.

In Fig. 1, probability density functions of the estimated distributions for all data sets and all candidate distributions are depicted in logarithmic scale in order to evaluate the tail characteristics of the fitted distributions. Additionally, in all the sub-figures, the data value range between -30 and 30 mVs is plotted with a zoomed view, to show the performance of the distributions around zero mVs. Examining all the sub-figures in Fig. 1, superior performance of the $S\alpha S$ distribution can be clearly seen in all the sub-figures. For noise amplitudes around 0, all five estimated distributions approximately coincide; however, for the tails, estimates for BG, GG and Middleton distributions are very poor compared to $S\alpha S$ and Student’s $t$. Specifically, the BG distribution tries to model the tail behaviour (especially visible for Lab2 data) thanks to its independent Gaussian parameters. Nonetheless, this effort is not successful as described by the performance measures given in Table 2. The performance of $S\alpha S$ and Student’s $t$ are very close for Lab2; however, for the rest of the measurements, $S\alpha S$ provides the best fit among all distributions.

4 Conclusions

In this paper, we performed a comparison study in modelling the impulsive noise for indoor PLC systems among five candidate statistical models. The results demonstrated the superior performance of $S\alpha S$ distribution among the others and subsequently exhibited evidence on the suitability of $S\alpha S$ distribution in modelling PLC impulsive noise.

In the literature, there are studies which model the impulsive noise in PLC systems by $\alpha$-stable distributions. Particularly, these studies provide a direct modelling scheme via $\alpha$-stable distribution, whereas the study in this paper decided the best model among five impulsive distribution families in terms of statistical error measures. The results presented in this study contribute to the literature with an experimental verification to [8,13] by demonstrating the importance of $S\alpha S$ distributions in PLC impulsive noise modelling. Thus, in applications such as emulator software and noise cancellers which mainly assume a Middleton Class A distribution (or a BG model), we clearly demonstrated the need for using $S\alpha S$ distributions.

Despite its common usage, the results presented in this study conclude that the Middleton Class A model is not the most suitable choice and has a limited capacity to model impulsiveness in indoor PLC measurements by experimentally verifying [11]. Even though it is a simplified version of the Middleton Class A with two Gaussian components, the BG model further outperformed Middleton Class A for indoor PLC noise measurements, as in the outdoor cases reported in [11].

It is interesting to note that all the considered models are scale mixtures of Gaussian distributions [30–32]. In particular, Middleton Class A model has Poisson mixing distribution [11]. Recent work by Lemke et al. [33] also shows how the $\alpha$-stable random variables can be expressed in terms of Poisson random variables providing a very interesting link between Middleton Class A and $\alpha$-stable distributions. Besides, $\alpha$-stable distributions have a natural strength to model noise processes in any system with additive nature and any level of impulsiveness, since they satisfy the generalised central limit theorem [19]. They were proven also analytically to be natural noise models in communication systems in an earlier study [34].

Moreover, we conclude that the Student’s $t$ distribution exhibits a considerable modelling performance as being the second best model after $S\alpha S$. Considering that the Student’s $t$ distribution has never been used in PLC noise modelling, we believe that the results in this study make it an easier-to-implement an alternative model to $S\alpha S$ distributions in future studies.

Acknowledgements We would like to thank Dr. Paulo Alexandre Crisóstomo Lopes (Instituto Superior Técnico, Portugal), for providing us with the PLC noise measurements used in this study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Gungor, V.C., Sahin, D., Kocak, T., Ergut, S., Buccella, C., Cecati, C., Hancke, G.P.: Smart grid technologies: communication technologies and standards. IEEE Trans. Ind. Inform. 7(4), 529–539 (2011)

2. Balbuena-Campuzano, C.A., García-Ugalde, F.J.: Performance of HSR and QPP-based interleavers for turbo coding on power line communication systems. Signal Image Video Processing 8(4), 615–624 (2014)
