NB-PLC Frequency Bands (151-471 kHz) Feasibility Study for PRIME v1.4 Protocol

Arias Blanco, Victor
Master's Degree in Smart Grids
Universidad Pontificia Comillas
Madrid, Spain

Matanza Domingo, Javier
Telecomunications Department
Universidad Pontificia Comillas
Madrid, Spain

Sendín Escalona, Alberto
Global SmartGrids
IBERDROLA
Bilbao, Spain

Abstract — *Narrowband Power Line Communication (NB-PLC) is one of the most used technologies for smart metering access. With the development of these technologies, devices and protocols increase the available frequency bands. This is the case of the new version of the PRIME v1.4 protocol, whose the new frequency bands are the subject of study of this project.*

This article seeks to collect existing frequency spectrum information between 151 and 471 kHz and then compare it with its own field performance analysis. After the analysis, a grouping of the counters with similar communication patterns will be carried out, based on performance in each frequency band. After that grouping, common network topology properties are searched for each group.

Keywords — PRIME, PLC, FCC Band, Noise, Clustering, Classification, Smart Metering

I. INTRODUCTION

Smart meters represent the first step towards mass digitization of the low voltage network. This smart grid is not only referred as such by the possibility of transmission of consumptions by the meters, but also by the possibility of reading network variables, receiving connection/ disconnection orders, or the reduction of consumption, which will be necessary actions for advanced operation of the future grid.

The deployment of smart meters has been commonly associated with the installation of Power Line Communication (PLC) technologies for communication. This technology has been developed due to the emergence of objectives to be achieved or difficulties to be solved. In this process, the PRIME (PoweRline Intelligent Metering Evolution) protocol appeared for the standardization of this communication in several of the distribution companies.

Typically, in Europe the so-called CENELEC-A band has been the one used, which would include frequencies from 3 to 95 kHz reserved for electricity suppliers. A new frequency band would then be added to reach 148.5 kHz, but this would be intended for in-home communication. With the appearance of bandwidth needs and avoidance of potential interference, the inclusion of commonly used bands on other continents was the target of new advances in protocols.

An update to the PRIME [1] protocol was introduced to allow for better and faster data transmission, which includes the expansion of six additional frequency bands between 151 and 471 kHz. This addition greatly expanded communication possibilities by adding higher bandwidth and frequency bands to reduce transmission noise.

Table I: PRIME v1.4 frequency bands

Channel	Frequency Bands (kHz)
1	42 - 89
2	97 – 144
3	151 – 198
4	206 - 253
5	261 – 308
6	315 - 362
7	370 – 417
8	424 - 471

New applications for the LV grid in the chase of the smart grid, like the real-time remote metering operations, integration of distributed energy resources (DER) or the grid sensing, would guide into the increase on bandwidth in the LV communication section.

The new frequency bands, besides the possibility of increasing the bandwidth, pursue the avoidance of electromagnetic interference. Typically, standards are used in all the electrical and telecommunication equipment for capping the electromagnetic interferences. While these standards are being created and as an assistance, other frequency bands can be used for avoiding punctual interferences in the spectrum.

The objective of this paper is to analyze the viability of the new frequency bands for achieving the objectives previously mentioned and search for topological variables that can directly influence the communication performance in any channel.

The organization of the article content is as follows. First, the objective and motivation of the article is settled in section

II. In section III, the procedure used in the project is explained. The results obtained in that process are analyzed in section IV. Finally, the conclusions extracted are discussed in section number V.

II. Project Definition

This project faces the task of gathering existing knowledge of the frequency bands between 151 and 471 kHz and applying it to analyze the quality of communication on the devices of the network. Communication with these devices can operate with different quality on different channels, a designation given to the frequency bands of the protocol. Identifying quality parameters and helping to clarify the reasons for those differences between devices and channels are tasks to perform in this article.

The results of this paper will help to improve the PLC communications in the access to smart metering, since will be possible to analyze the best channels to do it. But also, to partially forecast the channels performance if the correlation between topological variables and performance indicators is achieved.

III. Project Development

A. *State-of-the-art research*

State-of-the-art research provides some important concepts to verify in subsequent analysis. The first is increase of attenuation with frequency, caused mostly by the increase in impedance. This would generate lower intensity signals in high channels when the lines are longer [2], [3].

The second concept is that there is a higher noise level for typical loads at the lower frequencies of the spectrum, although there are more localized ones that generate a more homogeneous noise [4], [5]. Finally, it is found that a spike of interference due to the environment on one of the channels can reduce or disable the channel that suffers it [6].

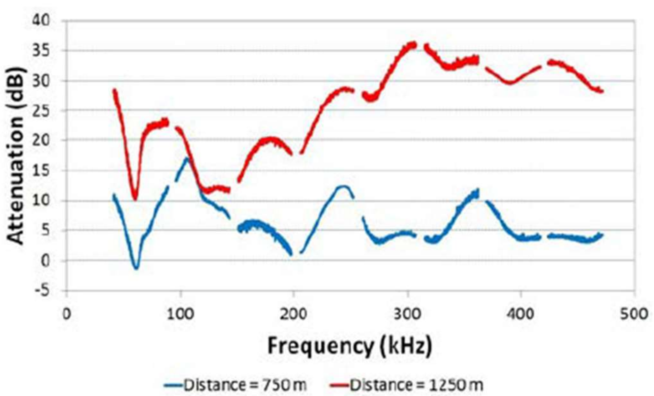

Fig. 1: Attenuation through frequency in a rural line

B. *Grid Topology*

The LV Grid begins in the Secondary Substation (SS) where the transformer lowers the voltage and connects to the Low Voltage Panel (LVP), one for each feeder. In the LVP is located a basic protection cabinet and connected to a remote management cabinet, where in just one of them is placed the Data Concentrator (DC). In this device is gathered all the data from the smart meters connected to this SS.

From the SS leave the feeders, mostly homogeneous and made of aluminum, varying on the cable diameter depending on the power. This homogeneity helps for generalizing the insights extracted from some locations.

The feeders reach the fuses boxes, the feeder protection cabinets, that will connect to the meters of the building. These meters are concentrated in a single space near the entrance.

From this research, it is acquired the necessary topological data for the future classification and also focus on the variables that can affect more drastically the PLC communication.

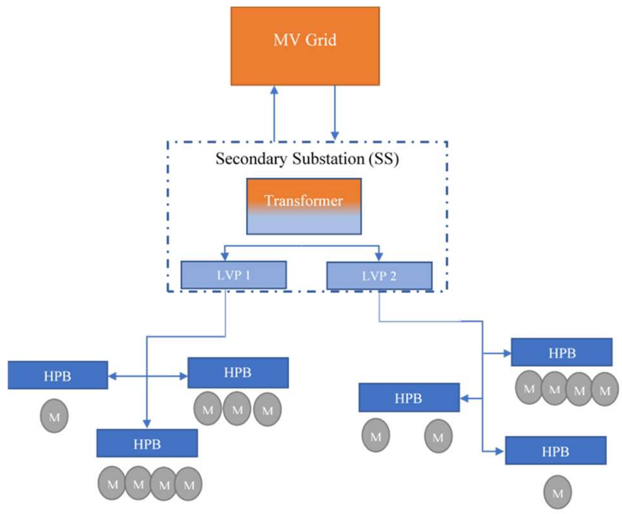

Fig. 2: LV Grid Topology Description

C. *Data Aquisition*

After analyzing the existing data, we proceed to acquire own data to make a performance study of these channels. Once the figures are processed, possible trends that can be highlighted are examined and possible performance indicators are extracted. Identifying the worst performing channels for each terminal allows them to be avoided, while optimal channels are prioritized.

The next step in development is to obtain topological characteristics that favor poor results in certain channels, to prevent from that risk in lines with similar conditions.

D. *Clustering*

The grouping process to identify poor communication quality on certain counters is performed by clustering with the *kmeans* method. The clustering is done mainly with two methods: the first is grouping depending on the performance pattern along the six channels studied, what implies the performance of a device in each channel; the second is getting the ones performing badly for each channel, independently of the performance in the others.

E. *Classification*

Once every device is tagged with the cluster it belongs, it is added to the topological data of the device. The classifier, that is a decision tree decided by try and error, try to identify which data can be used for emulating the division made by the clustering. The values of the variables used by the classifier can be obtained as topological characteristics of the group.

IV. Results

After a low voltage grid research, defining it in hierarchy and components and checking a high homogeneity in the materials, data from meters connected to processing centers was acquired in different locations. From these was extracted availability, Signal-to-Noise Ratio (SNR) and a signal strength indicator for each device and making an average every 10 minutes. With this information arranged for visualization, an uptrend was extracted from the median for availability, even though the SNR did not behave as favorably at the extremes of the spectrum. The signal strength indicator reduced its importance due to indirect calculation of the signal and dependence on noise, in which at low values of SNR, the signal strength value was unreliable.

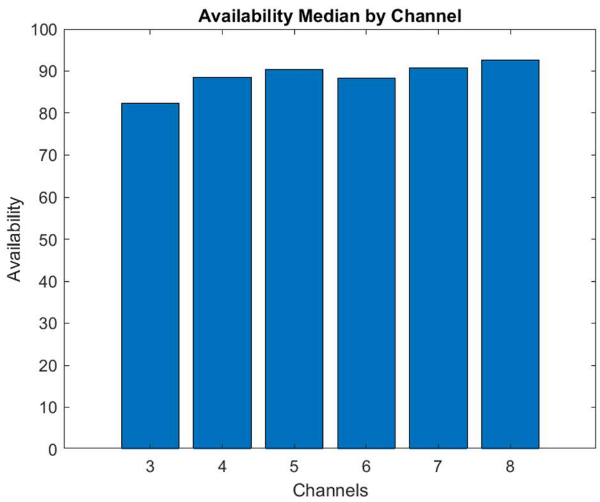

Fig. 3: Mean availability by channel along all the SS studied

After a calculation of daily statistical variables for each device, these were introduced in a variable reduction process. Those calculus could be the mean, the median or the standard deviation for each channel and each performance indicator. Because the number of variables was complex to represent and manipulate, in addition to the existence of a correlation between the variables, others were calculated that were linear combinations of the originals. The method for generating the new variables was the Principal Component Analysis (PCA). This method calculates the vectors were a bigger variance of the values is explained, called principal components, being the values of the vector the coefficients for each variable. These principal components reduce the number of variables to a pair without losing much information about the variance of the data, reaching more than 95% explained with only two variables.

A. Clustering

The new variables are required to group the terminals based on the performance in the communication for each channel. That performance is represented by the principal components in a general way, but the relation with original will be based on the coefficients shown in Fig. 4. Grouping is done by a clustering method named *kmeans,* both using the combination of all the channels and channel by channel.

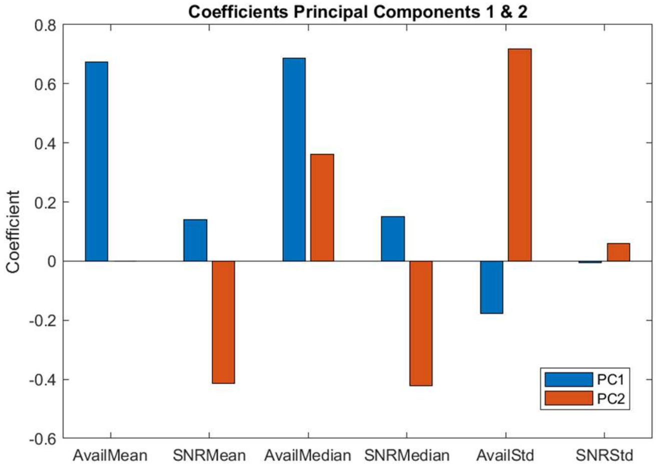

Fig. 4: Principal Components Coefficients

1) Channel combination clustering

When all the channels are used for the clustering, the groups found are patterns of performance along the six channels studied. Three groups were selected for dividing the data, due to not having clusters with too few elements.

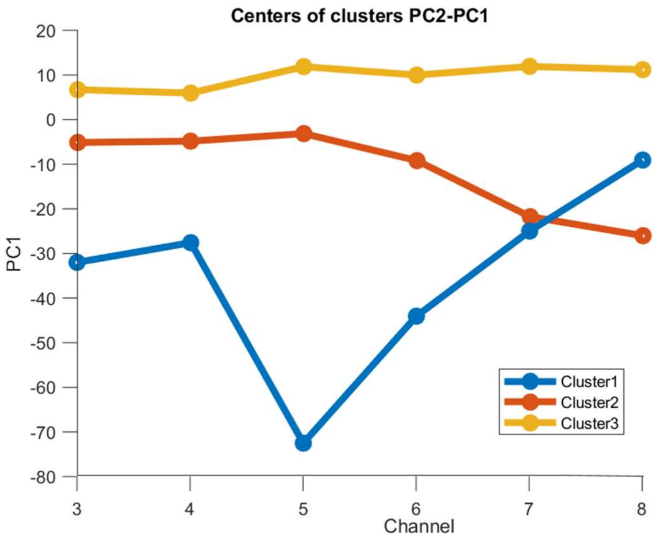

Fig. 5 Central value of the 1st component of the groups in each channel

Fig. 5 shows the score in the first principal component of the central values of the group for each channel. Cluster three shows good performance in every channel, while cluster two suffers a decrese in the upper ones. Last, cluster one has mostly bad values in middle channels and average in the rest.

Fig. 6 shows the median values of availability and SNR for the clusters seen in Fig. 5. Te principal component of the first figure is more related to availability, so the shape are similar although the values change. SNR do not correlate with availability for the upper channels, being clusters one and two the ones with best SNR but lower availability. The cause of this difference can lie on the effect of the attenuation on the SNR that does not have to be reflected on the availabillity.

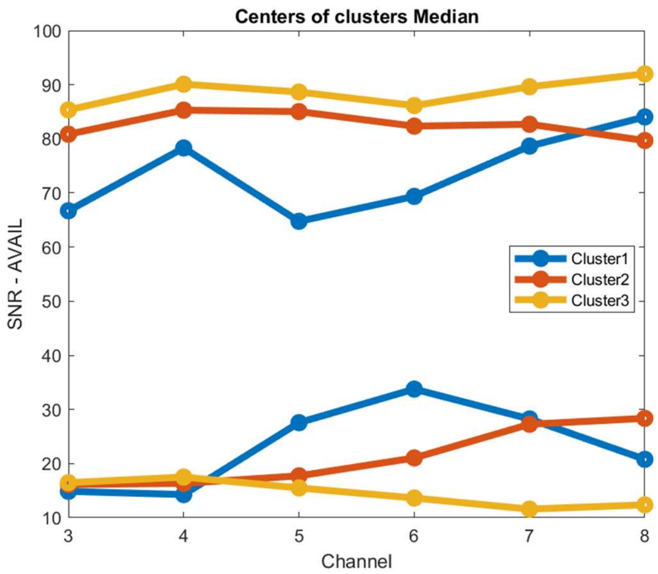

Fig. 6: Availability and SNR throughout channels for each cluster

2) *Channel by channel clustering*

When the clustering is done one channel at a time, the goal is to identify meters that have poor communication quality in that channel and assign a group tag to them. Fig. 7 shows the clustering according to the variables generated by linear combination of the originals. The image shown is for the seventh channel and the set is divided in two groups, one with acceptable performance and other with an inadmissible one. This conclusion can be drawn from these values because the coefficients of the linear combination are mostly positive, so high values of these variables are high values of the originals.

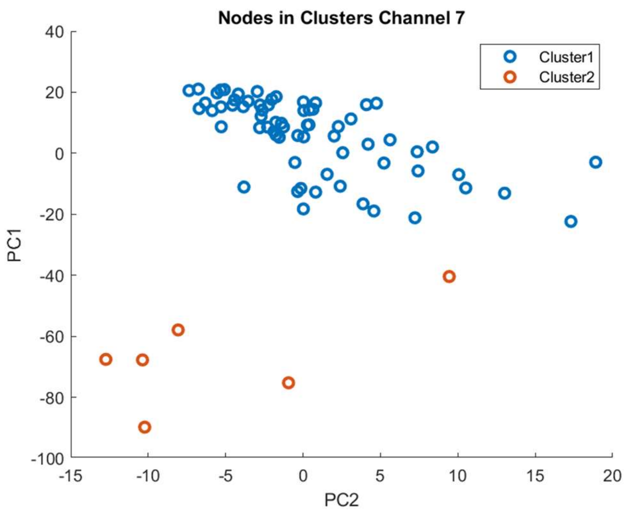

Fig. 7: Clustering terminals according to performance by main components in channel 7

Seventh channel is a good example of the clustering in the upper channels, with a clear division between the two groups. However, in lower channels the groups are not so well divided due to a higher correlation between the original variables used. Shows the total correlation between availability and SNR for median values in channel three, in this case original variables instead of principal components.

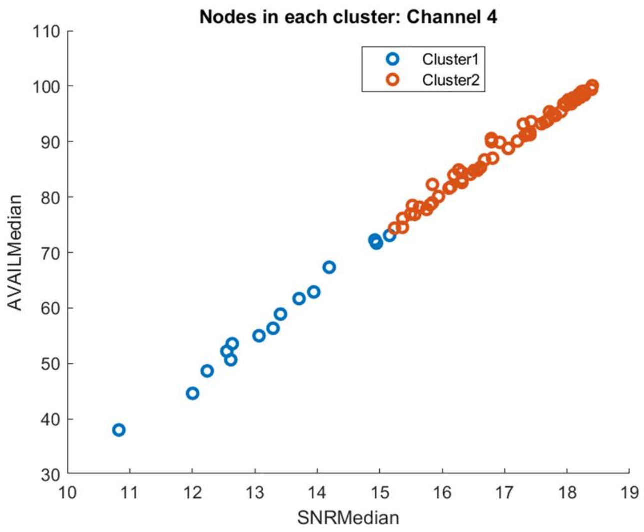

Fig. 8: Availability and SNR representation after clustering in channel 3

B. *Classification*

Once the devices have been grouped and tagged, a classification is performed with different variables. These new variables relate to the grid topology, such as the number of customers in the fuses box, the contracted power, or the distance to the secondary substation. If a decision tree, trained with some of the extracted data, is able to classify the terminals into the labels previously assigned with the topological variables, it implies that there is a correlation between the two sets of variables. If there is such a relationship between the performance variables in the communication and the selected topological variables, the decision tree will provide that values of the latter that generate a poor quality of communication.

1) *Channel combintaion classification*

Due to the difference in the elements of each cluster, the classification tree is biased into classify more inside the bigger cluster. Since this classification is more complex than a binary one, and uses data from different channels, more observations needs to be introduced for achieving a precise division. With the existent data, the classification is biased into the cluster with higher number of elements, but the variables with stronger importance in this classification have been the contracted power and the number of clients of the fuses box and the number of three-phase clients in the SS, as can be seen in Fig. 9.

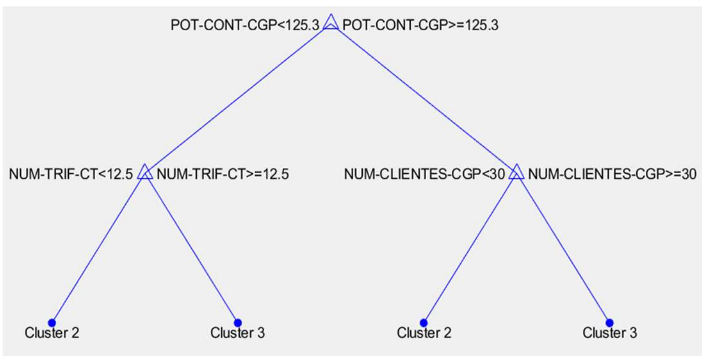

Fig. 9: Classification tree for channel combination

2) Channel by Channel clasification

Fig. 10 shows the classification test of a decision tree, previously trained with some of the data shown in Fig. 5. With the rest of the samples not used in the training, the effectiveness of the classifier was tested. In this case the result is very favorable just by erring a device, although the difference in the volume of samples per group can skew the classification.

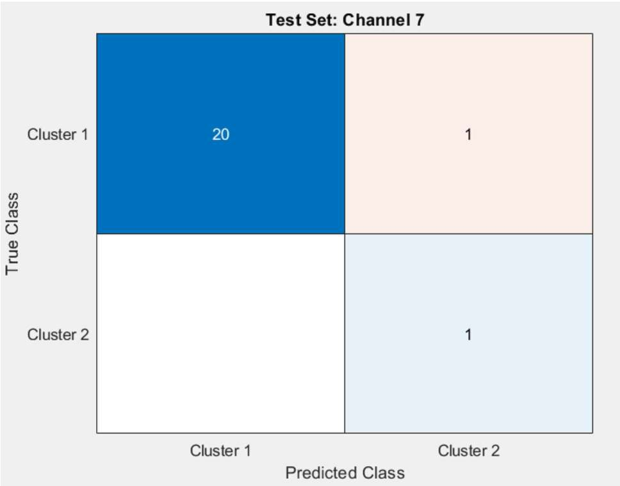

Fig. 10: Confusion matrix of test sample collection

The variables mainly used to achieve the separation between groups for channel seven were the distance to the transformation center and the type of line, whether underground or overhead. Bad data will be more likely on shorter lines or with overhead lines, according to the classification tree. In channels such as the third, divisions between groups are not clear which also prevents proper classification, for which, other variables should be included in both processes.

V. Conclusions

Nowadays there is a reduced amount of communication performance measures. In addition to this, it is complex to extract any conclusion from them without a deep knowledge of the PLC communication in LV grids and all the particularities of the grid where they are obtained.

A promising process of solving partially this situation is the PCA method combined with clustering. Through it, those measures can be combined and represented for then be divided depending on variables that consider all the information. This division, after a detailed adjust and iteration, can contribute to the analysis process of the practitioners finding tendencies and confirm or reject hypothesis. The attention to this method has to be maximum since from it is possible to do holistic analysis if it is adjusted properly, and all the future processes will be biased from the scores obtained from it.

From a stand-alone data analysis carried out, some of the conclusions have been extracted that could be generalized. First, the increased noise power in the lower channels and the higher signal attenuation in the upper ones seen in papers could be reflected in the SNR values of the measures. SNR is lower in third and eight channels while the middle ones have typically better values. These values are more differenced if the median is taken instead of the mean, reaching 3-5% bigger numbers. SNR is affected by the noise power and the signal strength, making low frequencies suffer from the presence of noise and high frequencies from the attenuation. Second, availability is not as much affected by the attenuation, this conclusion is due to the increasing availability with frequency, not reflecting the decrease seen in the SNR. An important characteristic to mention about the dataset is the decreasing availability-SNR correlation with frequency, since some of the worst availability values in the upper channels happens for the best SNR values, while in the lower ones both are completely correlated.

The clustering process has been predominantly successful and from it could be concluded two different aspects. The first one, from the clustering done combining channels, is the presence of at least three groups of nodes with common characteristics based on the performance along the six channels studied. It has been selected three clusters because they have enough nodes to be generalized with the existent data and gives useful information about the nodes included in each group. These three groups would be one with a good performance in every channel, a second one with better performance in the lower channels and a third one average on lower channels and good in the upper ones. The most populated one will be the first cluster.

From characterization done channel by channel, upper channels have a clearer distinction between good and bad performance. However, the total correlation between availability and SNR in lower frequency channels previously mentioned reduces the nodes variance and complicates the grouping. The measures obtained from the grid seems enough for a proper clustering of the upper channels, while others need to be found for the lower ones.

From the PCA analysis and the following clustering, a method for the identification of specific characteristics of the worst performers was consider useful. This would be a classification method, which performance is linked to the clustering. The classification data is fed from the clustering groups and topology data of those devices. With that knowledge, it tries to figure out the characteristics that divides each group. Lower channels are more inaccurately classified, also due to the imprecision of the previous clustering. Upper ones' classification is more accurate and seem to perform worse in longer lines, what is in accordance with the state-of-the-art conclusions.

Summing up the needs for improving the results observed in this work, the number of nodes analyzed is crucial for both clustering and classification since both uses data driven methods, especially for precision in the classification. Other improvements for each section would be: a bigger research on the set of variables introduced in the PCA, particularizing more the variables for the cases with a worse clustering result; adjust the number of clusters to the new data; the identification of variables that reflects indirectly better the noise or the common location can improve heavily the classification in the lower channels. Another future improvement in the process is the close work with technicians, to also work in the verification or rejection of stablished hypothesis instead of working from a blank page.

VI. ACKNOWLEDGEMENT

The author wants to thank Iberdrola for his support and contribution in this this study.

VII. REFERENCES

[1] PRIME Alliance TWG, «PRIME v1.4 White Paper». Accedido: may 06, 2020. [En línea]. Disponible en: https://www.prime-alliance.org/wp-content/uploads/2014/10/whitePaperPrimeV1p4_final.pdf.

[2] I. Fernandez, A. Arrinda, I. Angulo, D. De La Vega, N. Uribe-Perez, y A. Llano, «Field Trials for the Empirical Characterization of the Low Voltage Grid Access Impedance From 35 kHz to 500 kHz», IEEE Access, vol. 7, pp. 85786-85795, 2019, doi: 10.1109/ACCESS.2019.2924253.

[3] I. Fernández et al., «Characterization of the frequency-dependent transmission losses of the grid up to 500 kHz», Madrid, Spain, jun. 2019, vol. 1146.

[4] I. Fernández, D. de la Vega, A. Arrinda, I. Angulo, N. Uribe-Pérez, y A. Llano, «Field Trials for the Characterization of Non-Intentional Emissions at Low-Voltage Grid in the Frequency Range Assigned to NB-PLC Technologies», Electronics, vol. 8, n.o 9, Art. n.o 9, sep. 2019, doi: 10.3390/electronics8091044.

[5] I. Fernandez et al., «Characterization of non-intentional emissions from distributed energy resources up to 500 kHz: A case study in Spain», Int. J. Electr. Power Energy Syst., vol. 105, pp. 549-563, feb. 2019, doi: 10.1016/j.ijepes.2018.08.048.

[6] I. Arechalde, M. Castro, I. Garcia-Borreguero, A. Sendin, I. Urrutia, y A. Fernandez, «Performance of PLC communications in frequency bands from 150 kHz to 500 kHz», en 2017 IEEE International Symposium on Power Line Communications and its Applications (ISPLC), Madrid, Spain, 2017, pp. 1-5, doi: 10.1109/ISPLC.2017.7897123.