

# Application of smart meter phase connectivity to voltage unbalance in low voltage networks

**Author:** Garcerán Sánchez, Andrés  
*MSc Smart Grids Masters Degree*  
*Universidad Pontificia de Comillas*

**Academic Supervisor:** Matanza  
Domingo, Javier  
*Universidad Pontificia de Comillas*

**Industrial Supervisor:** Pazos  
Filgueira, Francisco José  
*i-DE (Iberdrola Group)*

**Abstract—** In this project the concepts of smart meter phase connectivity and advanced low voltage supervision are applications used for optimizing the connection of new customers to the low voltage grid. The methodology and model used in this project open the doors for creating a monitoring software of the LV network and optimizing new investment, future new connections and a smarter grid which can manage the future inclusion of the electric vehicles (EV) or demand response. With the equations used in the model, it is concluded that it is possible to simplify the calculation of intensities and voltages when the networks has a small percentage of three-phase loads, using coefficients developed during the project for simplifying the calculation from reactance of the cables, reactive measures and considering the models maximum error.

**Keywords—** Smart meter phase connectivity, Advanced low voltage supervision, Low voltage network, Voltage unbalance.

## I. INTRODUCTION

The most important new concepts brought from the smart grid investigations to this project are smart meter phase connectivity and advanced low voltage supervision.

First one of the concepts mentioned before consists in assigning the phase and the secondary substations to which each smart meter is connected too. Here is where big data and data science take place, with these tools by knowing the measures on the secondary substation and the measures on the smart meters using artificial intelligence techniques such as clustering [1] or using correlation techniques [2]. When using clustering techniques such as k-means clustering as in the model used in [2] an accuracy up to 90% can be achieved in clearing which is the phase in which each smart meter is connected to, as they say on their investigation depending on the month in which the data is picked as it is relevant for the output of their model. In the model they create 3 clusters and after they group the data, which phase is each cluster must be decided.

Advanced low voltage supervision has a main role as only secondary substations with this new feature would be considered. advanced low voltage supervision consists of a solution in which the main objective is to obtain more information from the power flows, the voltages, quality of the sinusoidal wave, non-technical losses, connectivity, fuses and neutral wires. For obtaining more information some sensors are installed connected to the neutral and phase wires and the secondary substation send the information through PLC technology to substation upwards which collects the information.

Thanks to big data techniques explain previously, the data of the phase and secondary substation to which each smart meter is connected has been obtained previously to this project. The project proposal is to develop a model that

calculates the intensity and voltage in the three phases and the neutral, as well as to get an image of the network topology and trying to optimize the information which has been used to adjust new clients.

The biggest concerns for the model and the project are dealing with are possible errors in the connectivity of the smart meters, possible non-technical losses and three-phase or not assigned-phase loads.

Three-phase and not assigned-phase loads have its importance cause these loads can change the distribution of power on each phase depending on the time of the day, week or year, the same load can have the biggest peak on phase 1 in the morning due to the nature of the industry or the consumption like for example, motor starting currents which happen in the beginning of the morning, and at the same time have the highest power consumption in phase 2 on the afternoon in winter do to the luminaires. Modelling three-phase and not assigned phase loads can be crucial for minimizing the error of the model estimations.

## II. CASE STUDIES

Four case studies were selected for the purpose of this project, of which three belong to the same secondary substation (A2, A5 and A6) and a fourth to a different secondary substation (B1). One of the most important characteristics of the networks, can be seen in TABLE I, which are the percentage of three-phase and not assigned-phase loads. A2, A6 and B1 have a small percentage of three-phase networks (less than 5%) compared to A5 which has a 25%. A6 was chosen for having a higher percentage of not assigned-phase loads.

TABLE I. CHARACTERISTICS OF THE LINES UNDER STUDY.

Case study	N° Smart Meters	N° Buses	Three-phase loads (%)	No-phase loads (%)
A2	32	6	3.13	6.25
A5	16	3	25.00	6.25
A6	29	3	3.45	13.79
B1	32	65	0.00	0.00

## III. METHODOLOGY USED FOR EACH CASE STUDY

The creation of a model, which will be explained more in detail in section IV, will be the first step. A model which will estimate the current and voltages of each case study. The model will optimize the disposition if the three-phase and not assigned-phase loads for minimizing the error of the model.

Once the power flow has been calculated, the analysis of the data takes place:

- As the most important data for making an investment or the operation of the LV grid are the maximum currents on the network, so an analysis for the 5% biggest current would take part to see the mean square error of the model for each line.

- An error analysis for measuring the cosine of the 5% max currents to discern the importance of the reactive power.

- An error analysis for estimating the importance of the reactive impedance on the voltage calculations.

Once the three analyses have been carried out for the four case studies, the coefficients that would be used to simplify the calculation when trying to match a new client will be calculated.

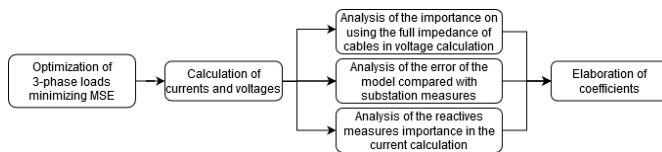


FIG. 1. METHODOLOGY USED FOR ANALYZING CASE STUDIES.

#### IV. MODEL

The model used and developed for the project was created by the author of this project and also the author of project [3]. It will be explained in next sections which parts were mainly done by the author of this project.

The first step for considering smart metering phase connectivity is to create a model of the LV network with the calculations required for making the comparison between the current measures of the advanced supervision on the secondary substation and to estimate the voltage on the consumers connection buses of the network.

The information used for the creation of the model was collected in excel files and MATLAB was used for extracting the information as the model was created in the MATLAB programming script tool. The excel files description used for this project can be seen below:

- For each secondary substation a file would have the same information: the topology of the network: type of cable used, end and beginning bus for each branch, the length of each branch, number of lines each secondary substation has and characteristics of the electrical protection box.

- Common to all secondary substations there is a file where the characteristics of the cables can be found and an excel file where the information of each smart meter about the line and secondary substation is located and the most important data of the project: the phase in which each smart meter whose has been through the big data algorithm has its phase assigned.

- For each electrical protection box: a file that collects the hourly power active and reactive consumption of each meter of the electrical protection box for a given period (typically 1 year).

- A file of the advanced low voltage supervisory measures for each line of the secondary substation: current of the three phases and the neutral, voltage of the three phases and the power exported and imported. This data was collected through a period of typically a year. An important fact about the advanced supervision is that data come within 5 minutes instead of hour consumption as smart meters, so the mean of those measures must be carried out.

- For each line there is an excel which connects the manufacturer code for each smart meter with the code number it has been assigned to each smart meter depending on the secondary substation they belong to.

The model will use the previous information and as a final goal estimate the current and voltage of the line in all buses and branches. The model will work for the information of one line of the secondary substation, not all lines of it. The model can be divided into the next parts:

1. Acquire the network topology: the first step consists of using the information recovered previously to have an image of the LV network for each case.

2. Compilation of the impedance of the cables of the network, phase wires and neutral and assign them to their respective branch

3. Assignment of each electrical protection box to the respective bus where it belongs, and the power consumptions of the smart meters are divided into their respective phase.

4. Calculation of the power flows in the network using an iterative calculation method and applying overlap techniques.

5. Last step is an optional step just if there are three phases or non-phase assigned consumers in which an optimization is realized to achieve the optimal three phase distribution compared to the secondary substation. Once the optimal three-phase consumers distribution has been achieved, step 4 is repeated.

##### A. Topology

In the data which was provided for this project for each branch it was included the origin bus and the final bus of each branch. An example of an image of the topology of the network can be seen in Fig. 2.

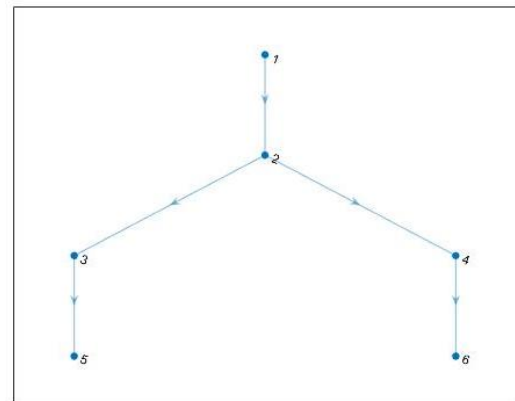


FIG. 2. EXAMPLE OF THE TOPOLOGY OF A NETWORK. SOURCE: SELF-MADE.

##### B. Optimization

For minimizing the error, the model could incur due to three-phase loads, an optimization algorithm was used for allocating three-phase and not assigned-phase loads to each phase minimizing the error. The algorithm consisted in allocating a patron, which consisted in a percentage of the power for each phase, shared for three-phase and not assigned-phase loads.

Two strategies were chosen: minimizing the mean square error of all phases of the 95<sup>th</sup> percentile maximum

currents or to minimize the mean square error of the neutral wire 95<sup>th</sup> percentile maximum currents. Formulas used for calculating the mean square error of the current are the next ones:

$$MSE_{I-phase} = \frac{\sum(I_{SS} - I_{Model})^2}{n_{phase}} \quad (1)$$

$$MSE_{TOT} = \frac{1}{n_{TOT}} (MSE_R * n_R + MSE_S * n_S + MSE_T * n_T + MSE_{neutral} * n_{neutral}) \quad (2)$$

The results of the optimization algorithm can be seen in TABLE II:

TABLE II. OPTIMIZATION RESUME OF THE CASE STUDIES.

Case study	Algorithm used	Phase R (%)	Phase S (%)	Phase T (%)
A2	OPT-TOT	48.11	43.19	8.71
	OPT-N	0	0	1
A5	OPT-TOT	25.27	21.69	53.04
	OPT-N	29.03	13.13	57.85
A6	OPT-TOT	51.30	0	48.70
	OPT-N	75.62	12.49	11.89
B1	-	-	-	-

### C. Current calculation

Once the topology and every branch has been assigned its correct impedance a function which was called createCurrentFile created from the data smart meters had for each hour of each day a unique matrix with three phases. For example, if the network has 36 users it means that there are 36 measures for each hour, this function assigns each electrical protection box (and the smart meters inside of it) to the corresponding bus and then allocates each smart meter on their respective phase and as a result there is just a matrix with 3 phases instead of 36 columns and as long as the period of the consumptions has been taken.

The strategy followed for trying to reduce the computational complexity of the model was to apply overlap techniques. A conventional overlap way of solving the problem was to overlap the power injected from all buses (if there are 6 buses, the overlap of 6 superpositions calculations).

In this project the technique used was to connect the initial bus (there is assumed there is just one initial bus) with all the final buses, as it can be seen in Fig. 2 the network is divided in two *subnetworks* which include buses 1,2,3 and 5 and 1,2,4 and 6. The main challenge of these overlap technique is that bus 2 appears in both *subnetworks* which will double the power injected by that bus. This issue is solved by adding a binary matrix in which if the bus power injected has already been included by another subnetwork, it doesn't add to this one.

The main advantage of these difference on the overlap is the reduction from 6 calculations of the different buses to 2, as in this technique explained you just need to know the *subnetworks*. A very important advantage as the function will

be used as many times as the iteration of the voltages occur and can decrease the time needed.

In order to be able to calculate the voltages and the current as accurate as it is possible with the data available a function (called *currentClaculation\_Iter*) was created on the code of the model to include the current calculation inside of the voltage calculation iteration.

As the network is assumed it is unbalanced, the calculation can't be done using three phase equations as they assume network is balanced. Equations used will be single phased and all the three phases will be treated on different equations.

For calculating the current through the cables, the equations used are:

$$\vec{I}_{injected,phase} = \frac{\vec{S}_{meter,phase}}{\vec{V}_{Bus}} \quad (3)$$

$$\vec{I}_{branch} = B_{injected} * \vec{I}_{injected} + \sum \vec{I}_{previous-buses} \quad (4)$$

$$\vec{V}_{SS(R,S,T)} = cte \quad (5)$$

$$\vec{I}_{neutral} = \vec{I}_R + \vec{I}_S + \vec{I}_T \quad (6)$$

The first time the voltage calculation code runned, in the first iteration equation (3) the  $V_{bus}$  value is the voltage measured on the secondary substation, but once the iteration code is running the voltage used is the voltage of the bus output of the previous calculation.

$B_{injected}$  is a binary matrix used in equation (4) has been added for not including twice the same power consumption of the bus in two different *subnetworks*.

The current calculation starts from the bottom of the networks as the first current known are the last buses of the network, which means that the term  $I_{previous-buses}$  refers to the buses which are downwards compared to the one is been calculated.

### D. Voltage calculation

The voltage calculation iteration method was created for other projects [3] and in this project this method will be explained for understanding the output needed.

The iteration method consists in calculating the currents with the function and the equations explained in section C and once the current calculation is done the voltage estimation takes places with the following equations:

$$\vec{V}_{bus} = \vec{V}_{previous-bus} + \vec{I}_{branch} * \vec{Z}_{branch} \quad (7)$$

$$\vec{V}_{bus} = \vec{V}_{previous-bus} + \vec{I}_{branch} * \vec{Z}_{branch} \quad (8)$$

## V. RESULTS

### A. 95<sup>th</sup> percentile error study results

In this section an analyse of the performance of the model on the 5% highest peaks of current compared to the measures

of the secondary substation will be carried out. In TABLE III a resume of the results of the study can be seen.

TABLE III. RESUME OF 95<sup>TH</sup> PERCENTILE ERROR STUDY.

Case study	Optimization data	Phase	Mean Error per phase (%)	Mean of the Error of Phases (%)
A2	OPT-TOT	R	-	4.87
		S	2.49	
		T	6.65	
		N	31.71	
	OPT-N	R	-	8.32
		S	4.53	
		T	10.21	
		N	27.26	
A5	OPT-TOT	R	6.95	10.31
		S	12.38	
		T	9.66	
		N	22.98	
	OPT-N	R	20.59	14.66
		S	29.82	
		T	12.61	
		N	15.22	
A6	OPT-TOT	R	12.90	6.81
		S	-	
		T	4.24	
		N	16.39	
	OPT-N	R	14.93	6.97
		S	-	
		T	4.65	
		N	8.87	
B1	OPT-TOT	R	6.40	6.60
		S	-	
		T	-	
		N	7.35	

An important detail is about how the 95<sup>th</sup> percentile has been chosen taking into account all phases considering too the neutral, which means that if the line has a huge unbalance in one phase, for example in phase R, as R has higher currents on the 5% percentile would have more current of the R phase than other and maybe other simply don't have any point due to never having significant peak loads. An example can be seen in B1 case, as there are only phase R and neutral currents values in the 95<sup>th</sup> percentile.

As it can be seen neutral errors are lower for the three-phase distribution of the optimal neutral error. On taking the same line, for example A2, the mean error of the phases is always lower in the three-phase distribution of the optimal total error compared to the distribution for the neutral error. The highest error for the OPT-TOT distribution is for A5 line

with an 10.31% mean error, as A5 is the line with higher percentage of three-phase loads it is the line more affected about the optimization result and as it was explained in the introduction of the document three-phase loads can have the same power consumption with different distribution of the power.

The best mean error of the phases belongs to B1 the only line without three-phase loads or loads with no phase assigned.

As considering the neutral optimal error doesn't has a significant impact in neutral cable error but increases the error on the three phase it won't be considering for the conclusions. The TABLE IV which shows the results just has the results for the total error optimal algorithm used.

TABLE IV. RESUME OF 95<sup>TH</sup> PERCENTILE ERROR STUDY FOR TOTAL ERROR OPTIMAL ALGORITHM ESTIMATIONS.

Case study	Mean of the Error of Phases (%)	Max Error in phases (%)	Mean of the Error on the neutral (%)	Max Error in neutral phase (%)
A2	4.87	10.8	31.71	41.2
A5	10.31	58.6	27.26	67.4
A6	6.81	22.4	16.39	26.7
B1	6.6	14.6	7.35	17.3

From the analysis of the error the model performed on the 5<sup>th</sup> percent peak the additional error the three-phase loads add to the model is much more significant compared to the load which phase was not provided but the model for having a minimum error the line must have only one-phase lines. The model on B1 (a line with only one-phase loads) in TABLE IV it is shown a 6.60% of error while A5 (the line with most three-phase loads) has an 10.31% mean error. A2 and A6 have a considered number of loads without any phase assigned but these loads are less significant to the model as the reason must haven't been assigned is due to their lower consumption making in the end a smaller impact in the current.

Some of the issues which the model had to deal with were wrong assignments of the connectivity of the smart meters as a wrong assignment of the phase could lead to a big error. Another issue the model had to deal with is the fact that the supervised measures of the secondary substations are taken each 5 minutes while the measures of the smart meters are an average over one hour. Which could lead to values which one of the devices captures and the other doesn't. Last big issue which could appear are fraud connections, as a big error in one single phase could happen one possible explanation would be fraud and the only way to assure it is by going physically to the cable.

Limitation of data is a big concern for the model as possible fraud, mistakes in the phase or secondary substation assignment could lead to big errors on the respective phase and mostly the neutral wire which is the more sensitive to errors. More case studies could help to minimize the impact of fraud and wrong phase connectivity models.

### B. Results of the analysis on the reactance of cables impact on voltage drop

The main objective for analysing the difference on the voltage drop calculation between considering the full impedance of the cables of the line or just the resistance. In TABLE V the results of the analysis can be seen, all error including the phases and the neutral phase are all similar as cables used in the lines have similar characteristics of R/X. A5 and A2 lines just have 2 branches and as the same current is circulating on them the cables are the same, and neutral in these two cases are the same to the phase cables which makes it the same error.

TABLE V. ANALYSIS ON THE REACTANCE OF CABLES IMPACT ON VOLTAGE DROP RESULTS.

Case study	Mean of the Error of Phases (%)	Mean of the Error on the neutral (%)
A2	5.43	5.36
A5	5.64	5.65
A6	5.13	5.13
B1	5.83	5.88

All errors are on values from 5 to 6%, both on the phases and the neutral which means the R/X relation would be near 3, as it can be seen in equation (9). If a 6% is an acceptable error, you can consider the resistance of the wires instead of the whole impedance when R/X has a relation of 3 or higher.

$$Error \approx 0.06 = \frac{\sqrt{\left(\frac{R}{X}\right)^2 + 1^2} - \frac{R}{X}}{\sqrt{\left(\frac{R}{X}\right)^2 + 1^2}} \rightarrow \frac{R}{X} \approx 3 \quad (9)$$

### C. Results of the analysis of the phase cosine of the consumptions of the 95<sup>th</sup> percentile

Phase cosine of the power consumptions (smart meters measurements) of the 5% peak consumption is analysed for the same reason the impedance of the cables was, to see the impact of reactive current in the calculation of the model.

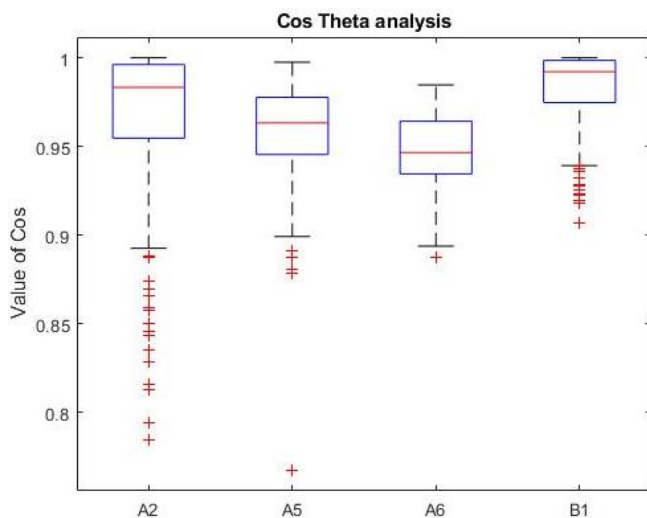


FIG. 3 BOXPLOT OF THE ANALYSIS OF THE PHASE COSINE OF THE POWER CONSUMPTIONS. SOURCE: SELF-MADE.

TABLE VI and Fig. 3 shows the results of the analysis, in all case studies cosine values oscillate between 0.9 and 1

having some outliers data below the 0.9 on A2, A5 and A6. Cosine values of the highest peak consumptions, in this case 5% peak loads, have statistically higher cosine phase values than low consumptions which means that reactive measures are more representative for low consumptions, which creates an opportunity of simplifying the calculations in high-load scenarios without the reactive values.

TABLE VI. ANALYSIS OF THE PHASE COSINE OF THE POWER CONSUMPTIONS.

Case study	Average cosine	Mean of the average cosines
A2	0.969	0.962
A5	0.958	
A6	0.946	
B1	0.986	

Having an average cosine of the measurements of 0.962 means the relation P/Q is 3.5, so the difference on the size of active and reactive measurements is clear. In equation (11) it can be seen the error made by not using reactive power is up to a 3.85%.

$$\cos \phi \approx 0.962 = \frac{\frac{P}{Q}}{\sqrt{\left(\frac{P}{Q}\right)^2 + 1^2}} \rightarrow \frac{P}{Q} \approx 3.5 \quad (10)$$

$$Error = \frac{\sqrt{(3.5)^2 + 1^2} - 3.5}{\sqrt{(3.5)^2 + 1^2}} * 100 = 3.85\% \quad (11)$$

## VI. CONCLUSIONS

In this project due to the multiple case studies and analysis which took place there are a variety of conclusions to be mentioned. There would be four mayor conclusions explained in this chapter.

First conclusion is not related to the objectives of the project and was an unexpected feature which appeared in the line A6. As it can be seen in Fig. 4 where it is shown the load curve of the phase S in amperes of a random day of the database comparing the secondary substation with both the optimal for the neutral error and for the total error distribution of three phase loads. As it can be seen measures of the centre are clearly on higher values than the model's output. An exhaustive look at the measures took place looking for a mistake in the model calculations and once it was sure the model was doing the calculations correctly, three hypotheses appeared:

- The possibility of having some meters which aren't in the list of meters of the line A6.
- A potential illegal connection can be occurring in the phase S.
- The measurement device of phase S of the secondary substation could not work at the efficiency it should.

The conclusion from this hypothesis is the power tool which smart phase connectivity can add to detect more fraud, to correct misinformation in the database or a possible malfunction measure device.

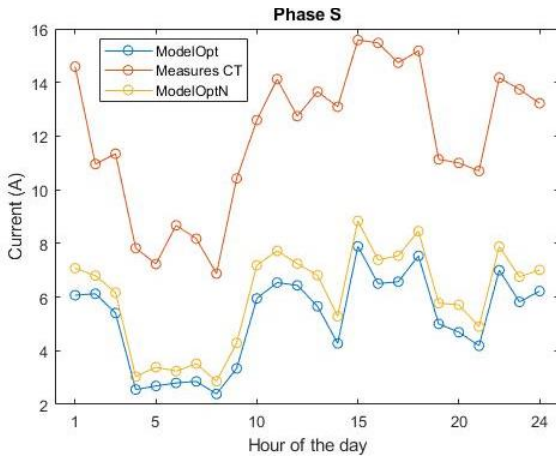


FIG. 4 PHASE S CURRENT CURVE OF A DAY A6. SOURCE: SELF-MADE.

From the analysis of the error the model performed on the 5<sup>th</sup> percent peak, it was proven the total error minimization lead to fewer errors and was chosen as the best model. It can be seen in TABLE VII a summary of the maximum errors of the case studies and the coefficient of safety taken from the calculation will be considered.

TABLE VII. MAXIMUM ERRORS OF THE MODEL FOR THE CASE STUDIES.

Case study	Mean of the Error of Phases (%)	Mean of the Error on the neutral (%)
A2	5.43	5.36
A5	5.64	5.65
A6	5.13	5.13
B1	5.83	5.88

As it can be seen in TABLE VII, case studies with low three-phase percentage such as A2, B1 and A6 (lower or equal to 3.63%) have less error than A5 which has a 25% of three-phase loads. Lines with high percentage of three-phase loads would need a different optimization algorithm to be more efficient and not exceed error of the 25%

For case studies with low three-phase percentage such as A2, B1 and A6 neutral error could go up to 41% which is a considerable error and would need a higher security coefficient than the phase which maximum is 22.4%. The coefficients would be:

$$C_{phase} = \frac{1}{1 - 0.224} \cong 1.30 \quad (12)$$

$$C_{neutral} = \frac{1}{1 - 0.41} \cong 1.70 \quad (13)$$

When analysing the result of the importance of the reactance of the cables the error which is always below the value of the calculation with the full impedance is 6% and it could be multiplied by a coefficient to minimize the error as it can be seen in equation (14).

$$C_{Reac} = \frac{1}{0.94} \cong 1.05 \quad (14)$$

Results of the importance of the reactive power of the loads consumes a mean error of 3.85% it is a small error which in case of trying to avoid the bias of just taking into account for the calculation the active power a coefficient could be multiplying the module of the power. The coefficient could be:

$$C_Q = \frac{1}{0.965} \cong 1.05 \quad (15)$$

As a summary of the conclusions which were described in this chapter the utilization of this coefficient would follow the next indications:

- Calculation of the currents and voltages using the equations of section IV
- If reactive measures are not used  $C_Q$  must be used multiplying the currents of the model to take considered the negative error which would appear.
- Voltage drop should get multiplied by  $C_{Reac}$  in case the reactance of the wires is not been considered.
- Depending on the percentage of three-phase loads if there is near to 4% or lower the security coefficients  $C_{Phase}$  and  $C_{Neutral}$  would be used for multiplying the current.

The use of the previous coefficient using the indications above could be the next one:

A consume with a  $P_{consumer}$  is a candidate for connecting to a line, a software with the model equations added up would calculate the current ( $i_{Line-phase}$ ) and multiply by the coefficient considered.

With the current estimated the voltage drop is calculated without considering the reactance of the wires and multiplied by the respective coefficient.

$$I_{Line-phase} = i_{Line-phase} * C_{phase} * C_Q \quad (16)$$

$$Capacity \geq I_{Line-phase} + \frac{P_{consumer}}{V_{SS} + \Delta V_{drop} * C_{Reac}} \quad (17)$$

The same calculation would be to see if the neutral wire has capacity for the consumer joining a phase of it is the limitation factor.

If the capacity of the phase wire or the neutral wire is not big enough the consumer would have to get a new line built or to be connected to a different line.

## REFERENCES

- [1] W. Wenyu, Y. Nanpeng, F. Brandon, D. Joshua and L. Juan, "Phase Identification in Electric Power Distribution," *15th IEEE International Conference on Machine Learning and Applications*, 2016.
- [2] L. Wenpeng, P. Joshua, M. Mirjana, L. Joyce and H. Brian, "Smart Meter Data Analytics for Distribution," *IEEE TRANSACTIONS ON SMART GRID, VOL. 6.*, 2015.
- [3] P. Gallego Fernández, "Use of current and energy reports to manage grid capacity and flexibility," 2020.