Faculdade de Engenharia da Universidade do Porto

Sensory fusion applied to power system state estimation considering information theory concepts

Bruna Tavares

Final Version

Dissertation conducted under the
Integrated Master in Electrical and Computers Engineering
Major Energy

Supervisor: Professor Doutor Vladimiro Henrique Barrosa Pinto de Miranda
Second Supervisor: Professor Doutor Jorge Pereira

26/7/2016

Resumo

Com o aparecimento de sincrofasores e com o crescente interesse na sua integração nas redes elétricas torna-se imperioso o desenvolvimento de métodos capazes de fundir a informação obtida através de diferentes classes de sensores, nomeadamente, informação obtida através de sensores convencionais integrados no sistema SCADA e sincrofasores, também conhecidos como Phase Meaurement Units - PMU.

Assim sendo, esta tese apresenta uma proposta de fusão sensorial que permite a integração de diferentes classes de sensores na estimação de estado podendo ser dada mais confiança a uma classe ou a outra.

Para além da fusão sensorial, este trabalho foca-se também na integração de conceitos relacionados com teoria de informação à estimação de estado, propondo uma alternativa à convencional estimação de estado resolvida através de mínimos quadrados. A substituição do critério de mínimos quadrados por conceitos relacionados com teoria de informação pretende propor uma nova vertente para um estimador de estado robusto sem necessidade de pré-tratamento da informação recolhida pelos sensores.

Abstract

With the increasing interest on the integration of synchronized phasor measurement units (PMU) in electric network appears the necessity to create methods capable to merge the information obtained from different classes of sensors, namely, the conventional sensors already integrated in the SCADA system and synchronized phasor measurements units, also known as PMU.

Thus, this dissertation proposes a sensory fusion method that guarantees the previous requirement, allowing the integration of different classes of sensors for the state estimation and assignment of different level of trust to each sensory system.

Beyond that, this thesis focuses on the appliance of information theory related concepts on the state estimation, proposing an alternative to the conventional state estimation solved by least square error. The substitution of the least square criterion of the state estimator by information theory concepts aimed at proposing a robust state estimator without the necessity of previous treatment of the acquired data.

Key words: Correntropy , Information theory, Parzen window, Phasor Measurement Units, Renyi's quadratic entropy, sensory fusion, state estimation, weighted-least square error.

Acknowledgments

First of all, I would like to thank my thesis advisor Professor Vladimiro Miranda of the Electrical and Computer Engineering Department at Porto's University. From the great and innovative ideas to the urge motivation, Prof. Miranda was always a sustainer of this work. Without his support and guidance, this thesis wouldn't have followed such a great path. I also want to address an acknowledgment to my co-supervisor, Professor Jorge Pereira. The door to Prof. Pereira's office was always open whenever I needed some help with challenges on my research or with problem resolution.

I want also to thank INESC Porto for the good reception and opportunity to integrate the CPES department. A special acknowledgment to CPES team, for all the support and good moments inside and outside INESC Porto. I was not expecting such a great environment, support and friendship between all the members in a company team. That was a good surprise and help my daily work with a good mood.

Finally, I must express my gratitude to my family, specially to my parents, for all the unfailing support and continuous motivation through all my life and my study years. Without their support I would never be the person I am today.

“Pleasure in the job puts perfection in the work”

-Aristotle

List of contents

List of figures

List of tables

List of acronyms and symbols

List of acronyms

CIM	Correntropy Induced Metric
CS	Cauchy-Schwarz
CPES	Center for Power Energy Systems
DER	Distributed Energy Resources
EMS	Energy Management System
INESC TEC	*INstituto de Engenharia de Sistemas e Computadores- Tecnologia e Ciência*
LSE	Least Square Error
MAPE	Mean Absolute Percentage Error
MCC	Maximum Correntropy Criterion
MSE	Mean Squared Error
p.u.	Per Unit
pdf	Probability Density Function
PSS/E	Power Transmission System Planning Software
PW	Parzen Window
RQE	Renyi's Quadratic Entropy
SCADA	Supervisory Control And Data Acquisition
SE	State Estimation
WLS	Weighted Least Square

List of symbols

- σ Standard deviation, also used in this work as PW's size
- Θ Voltage phase
- α Trust assign to the measurements obtained by sensors from group one.
- β Trust assign to the measurements obtained by sensors from group two.

Chapter 1

Introduction

In this Chapter, a brief overview about the addressed problem will be presented. First, the motivation to the development of this thesis will be discussed, referring the importance and the need of PMU integration in the measurement system as well as the benefits of information theory related concepts applied to the state estimation problem. Then, the purpose of this thesis will be explained and the objectives presented. Finally, the organization of this thesis will be exposed.

1.1 - Motivation

There is a need of a real time monitoring and control system in the electrical power system. This function is traditionally performed by a SCADA, Supervisory Control And Data Acquisition, and a EMS, Energy Management System. In the existing networks, measurements are provided by conventional sensors (devices) that measure: line power flows, bus power injections and bus voltage magnitudes. Then, a state estimator is required, so that all of the systems' components estimations can be simultaneously possible taking into account the grid configuration and the electrical circuit laws. Thus, by correcting the original measurements values in a way that each corrected measurement (estimation) is coherent with all the others, the function of the state estimation is to find a valid operation point (vector). Despite little deviations requiring a state estimation procedure, there is also the possibility of contamination of the measurement set by gross errors. Such errors need special attention and a supplementary treatment.

Presently the criterion mostly used in power system state estimation is the minimum square error (also known as least squares). Due to the fact that the state estimation is done under a quadratic criterion, it has the inconvenient that the influence of bad data is also quadratic, which leads to a bad behavior of the estimator. This approach is, therefore, mostly useful if the subjacent error distribution is Gaussian. As the presence of gross errors may seriously distort such distributions, a preprocessing of the measured data in order to filter gross errors is required, albeit the present day techniques not being totally satisfactory. This thesis presents another approach based in information theory which can adequately deal with gross errors, making their presence mostly irrelevant. That approach was already proved for Correntropy [1] and tested. In this thesis, other information theory related metrics are tested.

Lately, the necessity of integration, in an estimation process, of very distinct types of sensors has become evident, because of the growing use of PMU, Phase Measurement Units. PMU have a lot of advantages over conventional sensors for modeling real-time systems. However, a total replacement of the SCADA system by a synchrophasor based system is inconceivable in the short-time: first, the number of PMU presently installed over the networks do not allow a full observability of the system; and second, to achieve a total observability would be necessary a big investment to replace material that is still in its useful "lifespan" by new PMU, not forgetting the big investment in infrastructures to support the complex SCADA system [2]. Therefore, the two sensory systems (conventional and PMU) are going to coexist along the coming years. Thereby, there is a need to develop efficient methods of giving coherency, in a single estimation process, to two sensory systems of very distinct nature. This process, of putting together distinct sensory systems in order to build up a coherent map of an external reality is denoted "sensory fusion" in the robotics area and will be adopted in this thesis.

Thus, the importance of developing an innovating work on sensory fusion becomes evident. Because building a map of an external reality (the estimation process) must harmonize the information content of the datasets collected by different sensor systems, it seemed promising to explore new information theory related concepts, namely adopting them as new criteria for the state estimator, in other to increase the efficiency of the monitoring system.

1.2 - Objectives

The main objective of this work is to develop a sensory fusion method for state estimation. Such method must allow the coexistence of conventional sensors and PMU in the same network and build an estimation from a cooperation of both sensory systems. As it is known, the synchronized phasor measurement units have been developed. Due to the numerous advantages that those components present, the interest in their application has been growing. Therefore, this work aims to propose a new sensory fusion method that allows the integration of measurements from both systems in the state estimation, allowing the user/estimation process to give more importance to one system or to the other.

It is important to mention that, despite the importance of the data acquisition units, their measurements treatment is still more relevant. For that reason, in this work a special attention to the state estimation process is given, proposing another perspective and opening doors to a path in a different direction, where the attention given to each sensory systems depends on its probability to lead the estimation to the right state.

In this work, it is also proposed that information theory concepts may work as an alternative to the conventional least square error (LSE) criterion, applied for a long time in state estimation. The target is to find different and efficient metrics able to integrate a robust state estimator. These metrics mustn't decrease the quality of the results already obtained with the LSE, and must also present a good behavior in the presence of bad data and real error distributions.

As it is known, the LSE is only optimal if one assumes the error distribution as a Gaussian distribution – and that is not always, or rarely is, true. In this way, entropy concepts will be compared with the LSE as the state estimation objective function, inferring their advantages and disadvantages. The main goal is to find an innovative and appropriated alternative metric that presents good behavior facing outliers.

1.3 - Thesis organization

This thesis is organized on seven main chapters. The first one aims to present the purpose of this thesis and its importance in real systems application, as well as a brief presentation of the innovator concepts applied to the state estimation.

The second chapter includes three main sections, presenting what is already developed in each concept related to that thesis. The first topic is information theory: concepts related with entropy that are going to be applied in the further chapters as alternative criteria to the least square error in state estimation process. The second is related with the state estimation problem, its formulation and new challenges that it has been facing. The third is about sensory data fusion, intending to explain its importance and the situations that should be avoided, such as catastrophic fusion.

The third chapter applies four different metrics as objective function of the state estimation, comparing their behaviors. The Least Square Error (the conventional and most utilized method), Correntropy, Renyi's Quadratic Entropy and Cauchy-Schwarz. These three last concepts are related to the information theory and are going to be explored on the third chapter, assuming that they may be applicable to the state estimation problem. The advantages of the applicable concepts in comparison with the conventional method are going to be analyzed and presented.

The fourth chapter discusses the possibility to separate the set of measurements of the system in two groups, according to the measurements origin (conventional sensor or PMU), and computes the estimation assigning different trust values to each group of measurements. This analysis is done in the framework of a multiple criteria modeling and Pareto-front discovery. The trust variation results in different values to the evaluation of each group's estimations, drawing a curve with axis: (fitness evaluation of group one, fitness evaluation of group two). The effect of the trust variation is analyzed to the different state evaluation metrics, LSE, MCC and RQE.

In the fifth chapter, some fusion metrics are proposed, applied and analyzed in order to find an optimal fusion point, over the trust variation curve, that approaches better or is closer to an ideal point. That ideal point is characterized by two components (x,y)=(maximum fitness evaluation of the group one, maximum fitness evaluation of the group two) and it's outside the trust variation curve, as it can be seen in chapter 4. This concept of ideal was disseminated by Milan Zeleny [3]

Many tests for the original proposals studied in these chapters were made adopting a DC model. It was, however, necessary to test the developed models in a realistic system. Therefore, in the sixth chapter the most interesting fusion metric and evaluation metrics are applied in an AC model of a network based in real European medium voltage system. In this chapter, some of the main points are:

- Illustrating that the introduction of PMU in the measurements systems allows better state estimation, decreasing the errors on the estimation;
- Illustrating that the MCC and RQE criteria applied to the state estimation allow the identification of bad data, such as missing measurements and inversion of line power flows, without pretreatment of the measurement values, creating a robust state estimator;
- Testing the fusion metric previously proposed in a realistic measurement system.

Finally, the main conclusions about the proposed state evaluation metrics and about the new sensory data fusion model are presented in chapter seven, together with some proposals for future work.

An annex is added to the thesis with some complementary information: the test system and the process used to generate measurement sets. A second annex is present with some results that help the analysis in chapter 5. Also a long abstract of this thesis can be found attached.

Chapter 2

State of the art

This chapter presents important aspects related with the three main pillars of this thesis, namely: Information theory, state estimation and sensory fusion. Each one of these topics are presented in different sections of this chapter. These sections not only expose the theory inherent to the respective topic but also some progresses in the area.

2.1 - Information Theoretic Learning

This section mentions the information theoretic learning origins as well as some different approaches and developments.

2.1.1 - Origin of Information theory

Communication systems were the information theory crib when Shannon faced the problem of optimally transmitting messages over noisy channels [4]. Despite Communication systems have a physical substrate (antennas, transmitters, receivers), the essence of information theory deals with the mathematical components of communication systems such as the characterization of message's structure and limitation of error-free transmission of the message's content. Although new mathematical concepts usually require a long time and a lot of reviews from different people to be accepted and applied to practical problems, information theory was planned by only one man: Claude Shannon. His theory and accepted in a short time by science and engineering communities. Thus, information theory had a prompt impact in the design of communication systems and it implemented a mathematical framework to formulate and characterize interaction beyond physical laws [5].

The main reason why information theory was created was to help study the theoretical issues of optimally codification for messages according to their statistical structure. Therefore, it characterizes transmission rates according to the noise levels in the channel and evaluating the minimal distortion in messages. For that task it is necessary only two statistical descriptors: entropy and mutual information. Mutual information is strongly related with divergence in probability spaces, behaving a dissimilarity measure.

Through Shannon's view, there are three fundamental steps to obtain the optimal messages transmission with a predetermined distortion over noisy channels: First, the compression. When

the message rate is too high for the channel, it is necessary to limit the massage rate. That limitation leads to a need of optimally messages compression. The compression is done as established by the rate distortion theory, minimizing the mutual information between the original message and its compressed version. After that, there is the phase of error correction. It's necessary to encode the source-compressed data for error-free transmission in order to withstand the channel noise in the transmission. This step is achieved by maximizing the mutual information between the source-compressed message and the received message, as established by the channel capacity theorem. The last step, consists on the decompression. It aims to decode and decompress the received message to recover the original message. Basically, first it minimizes redundancy for efficiency and then it adds redundancy to mitigate the noise effects in the channel. The point here is that the same statistical descriptor, mutual information, is specifying the two compromises for error-free communication: the data compression minimum limit (rate distortion) and the data transmission maximum limit (channel capacity) [5].

The mathematical formulation for Shannon's Entropy is explored in the next sub-section.

2.1.2 - Shannon's entropy

Let's assume a random variable X, with $x_i \epsilon IR^D$ and a random sample A with n pairs $\{x_i, p(x_i)\}$, assuming that the probability density function (pdf) of the variable x is described by $P=\{(x_i,p_i(x_i)),\ i=1,2\ldots n\}$ or simply $P=\{(x_i,p_i),\ i=1,2\ldots n\}$, the entropy or uncertainty through Shannon is given by :

$$H_1(x) = E[-\log(P)] = -\sum_{i=1}^{n} p_i log p_i \ , \qquad (2.1)$$

where $\sum_{i=1}^{n} p_i = 1$ and $p_i \geq 0$. Therefore, Shannon defined the uncertainty of the variable X as the sum across the set of uncertainty in each component weighted by the probability of each component. The uncertainty of each component is given by the $log p_i$.

For continuous pdf distributions we have:

$$H_1(x) = -\int p(x) \log\big(p(x)\big)\, dx \ . \qquad (2.2)$$

The Shannon's entropy is seen as a measure of uncertainly in the information content of the pdf p(x). If the logarithm basis is 2, it obtains the information quantity is measured in bits.

2.1.3 - Renyi's entropy

In the mid-1950s, Alfred Renyi introduced the parametric family of entropies as a mathematical generalization of Shannon's Entropy. Renyi's entropy appeared when he tried to find the most general class of information measure. That measure should allow to sum statistically independent systems and should be compatible with Kolmogorov's probability axioms [5].

For the same conditions of the previous subsection the Renyi's entropy is defined as a family of functions related to a parameter α as follows

$$H_\alpha = \frac{1}{1-\alpha} log \sum_{i=1}^{n} p_i^\alpha \ , \qquad (2.3)$$

with $\alpha>0$ and $\alpha\neq1$. When $\alpha=1$ the expression losses its meaning. However, it's possible to demonstrate that when $\alpha \to 1^- \ and \ \alpha \to 1^+$ H_α tends to H_1 -Shannon's entropy. Thus, $\lim_{\alpha\to1} H_\alpha = H_1$.

Therefore, it can be assumed that the Renyi's entropy converges to Shannon entropy bilaterally. Also, Shannon's entropy is between the possible values of Renyi's definition, with real parameters α and ß. One has $H_\alpha \geq H_1 \geq H_\beta, where$ 0<α<1<ß.

Comparing Shannon's and Renyi's entropies, equations (2.2) and (2.3), the main difference is the placement of the logarithm in the expression. While in Shannon's entropy the $\log(p_i)$ term is weighted, in Renyi's entropy the log is outside the term that involves the α power of the probability mass function.

The parameter α makes the Renyi's entropy much more flexible enabling some different measurements of dissimilarity within a given distribution. Usually, $H_\alpha(x)$ is called the spectrum of Renyi information, seen as a function of α, and its graphical plot is useful in statistical inference.

In [5] is described some important axioms, as:

1. Renyi's Entropy measure, $H_\alpha(p_1, p_2, \dots p_n)$, is a continuous function of all the probabilities p_k. In this way, a small change in the probability distribution results in a small variation of the entropy.
2. $H_\alpha(p_1, p_2, \dots p_n)$ is permutationally symmetric. In other words, if the order within the vector p changes, the entropy value remains the same. The permutation of any p_k in the distribution doesn't affect the uncertainty or disorder of the distribution and, thus, it should not affect the entropy;
3. H(1/n, …, 1/n) is a monotonic increasing function of n. In an equiprobable distribution, the increase number of choices leads to an increase of uncertainty or disorder, increasing also the entropy measure.
4. Recursivity. In a recursive function it should possible to observe the follow condition

$$H_\alpha(p_1, p_2, \dots, p_n) = H_\alpha(p_1 + p_2, p_3, \dots, p_n) + (p_1 + p_2)\, H_\alpha\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right), \quad (2.4)$$

   which means that the entropy of N outcomes can be expressed in terms of the entropy of N − 1 outcomes plus the weighted entropy of the combined two outcomes;
5. Additivity. Considering two independent probability distributions, p = $(p_1, p_2, \dots, p_n)$ and q= $(q_1, q_2, \dots, q_n)$, the joint probability distribution is denoted by p·q , the property $H_\alpha(p.q) = H_\alpha(p) + H_\alpha(q)$ is called additivity .

2.1.4 - Renyi's quadratic entropy

Renyi's quadratic entropy, α=2, gets a special attention by its practical applications. It is represented by the follow expression

$$H_2 = -log \sum_{i=1}^{n} p_i^2 \ . \quad (2.5)$$

For continuous pdf distributions we have:

$$H_1(x) = -\frac{1}{1-2}\int (p(x))^2 dx \ . \quad (2.6)$$

This particular Renyi's entropy has an interesting geometric interpretation. Assuming a discrete random variable with a probability density function in the form of $P=\{(x_i, p_i(x_i)), i=1,2\dots n\}$, that pdf can be seen as a singular point in a probability space of n dimensions, where the axis I correspond to the $p_i(x_i)$ of each x_i. That point, corresponding to the coordinates $p_i(x_i)$, is over the hyperplane given by:

$$\sum_{i=1}^{n} p_i = 1$$

Therefore, the metric H_2 is just the (negative) logarithm of the Euclidian distance between that point and the origin.

2.1.5 - Parzen window

The Parzen window method is a powerful tool to interact with some concepts of information theory. Most of the times we assume a pdf as "real", "exact" or "hidden but assumed real"; however, many times what we have is a discrete set of points which are taken as a representative sample of the assumed unknown pdf. To provide an analytic description for an estimate $\hat{f}$ of the unknown pdf that would explain the present data in a sample of n points $y_i \in R^M$ in a M-dimensional space, the Parzen windows method is quite useful. This method consists in apply a Kernel function centered on each sample point. If a Gaussian Kernel is used, the estimated pdf $\hat{f}$ is obtained by the summation of the individual contributions of the Kernel applied on every point, as

$$\hat{f_Y}(z) = \frac{1}{n}\sum_1^n G(z - y_i, \sigma^2 I) \ , \tag{2.7}$$

where $G(z - y_i, \sigma^2 I)$ is the Gaussian Kernel function and $\sigma^2 I$ is the covariance matrix . It's possible to prove that, when $\sigma \rightarrow 0$ and $n \rightarrow \infty$ the estimated pdf tends to the real pdf, $\hat{f} \rightarrow f$ [6].

It's easy to understand that the Parzen window's size, characterized by the standard deviation σ, conditions the shape of the estimate for the true pdf distribution. Thereby, a higher standard deviation provides a smoother form, while a low σ generates a more irregular pdf. This influence is possible to observe in the following charts, figures 2-1 to 2-3.

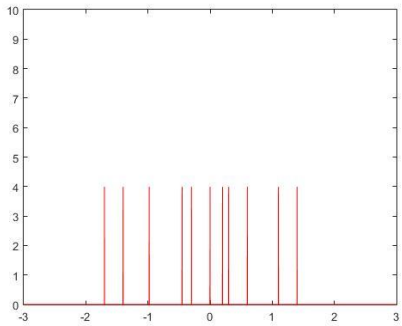

Figure 2-1 - Graphic illustration of a sample of 8 points D={-1.70 -1.40 -0.98 -0.45 -0.30 0 0.20 0.30 0.60 1.10 1.40}, represented by Dirac pulses at each location.

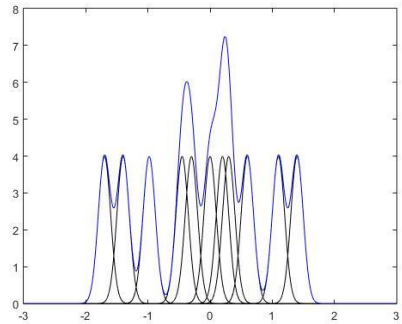

Figure 2-2 – pdf estimated by PW method for a set of 11 points, with σ=0.1. The blue line represents the sum of the Gaussians. The estimated pdf corresponds to the blue line divided by 11, in way that its integral is equal to 1.

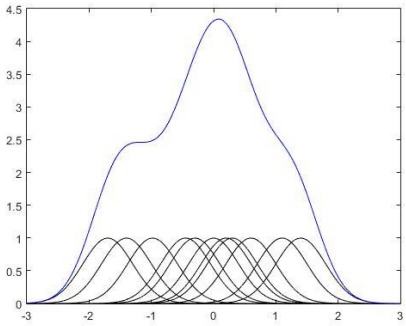

Figure 2-3 - pdf estimated by PW method for a set of 11 points, with σ=0.4. The blue line represents the sum of the Gaussians. The estimated pdf corresponds to the blue line divided by 11, in way that its integral is equal to 1.

As it is possible to observe for a higher value of σ, figure 2-3 the shape of the estimated pdf is smoother than the one provided by a lower standard deviation, figure 2-2.

In Chapter 3.1 -is possible to observe the application of the PW method to estimate an error distribution, based in the work [7]. The error distribution is need to the calculation of the RQE. However, only a discreet sample of points is known, the residual errors between the estimation and the measured values. Besides the estimation of the error distribution, the application of the PW method to the RQE concept also intends to introduce a parameter able to manage the expected error distribution. Thus, the proper definition of the parameter σ must allow the identification and natural elimination of large errors.

2.1.6 - Correntropy

Correntropy is related with the probability of two random variables being similar. Their similitude is measured in the neighborhood of the combined space controlled by a kernel function. That is, the window width defined by the kernel acts like a visor controlling the observation window where the similitude is evaluated. Actually, it can be said that Correntropy uses the Parzen window method controlled by the kernel function. This window has the effect of mitigate or eliminate the effect of outliers.

Correntropy is a proper measure for the similarity between two random variables X and Y, and can be expressed as

$$V_\sigma(X,Y) = E[G(X-Y,\sigma^2 I)] \ , \tag{2.8}$$

where the $G(X-Y,\sigma^2 I)$ is the Gaussian Kernel with variance σ^2 [8].

Truly speaking, the true pdf is unknown and only an estimative obtained from a data sample is possible. Thus, the estimated value of Correntropy can be calculated by

$$\hat{V}(X,Y) = \frac{1}{n}\sum_{j=1}^{n} G(x_i - y_i,\sigma^2 I) = \frac{1}{n}\sum_{j=1}^{n} G(\varepsilon_i,\sigma^2 I) \ , \tag{2.9}$$

where ε is the difference between the two variables.

Through [9] there are some important Correntropy properties, among which:

1. Correntropy is symmetric V(X,Y)=V(Y,X);
2. Correntropy is positive and bounded: 0<V(X,Y) ≤ $1/\sigma\sqrt{2\pi}$. The maximum $1/\sigma\sqrt{2\pi}$ is achived only when X=Y;
3. Correntropy involves all the even moments of the random variable ε=X-Y.

Applying the Taylor series development on the Gaussian Kernels it's possible to rewrite the Correntropy expression as

$$\hat{V}(X,Y) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{\sigma\sqrt{2\pi}}\sum_{k=0}^{\infty}\frac{(-1)^k}{(2\sigma^2)^k K!}E[(x-y)^{2k}] \ . \tag{2.10}$$

Correntropy considers all the even moments of the random variable ε. Also, its easily understood that the size σ of the kernel controls the impact of higher moments. When σ increases, the only relevant term becomes the second order moment, which is the expectation of the square difference between the variables- the variance of the error. Thus, it's fair to assume that Correntropy is a generalization of correlation, that is, it can be used as a measure of dependency between two random variables.
The advantage is that, by considering higher order moments, Correntropy can express non-linear dependences between random variables, and not only linear dependences. The same does not happen with correlation, where only second order moments are considered. Thus, in correlation, only linear dependences between random variables are considered.

4. Correntropy $\hat{V}_\sigma(X,Y)$ is the value of the pdf $\hat{f}_{\varepsilon,\sigma}(\varepsilon)$ evaluated at ε=0 ,

$$\hat{V}_\sigma(X,Y) = \hat{f}_{\varepsilon,\sigma}(0) \ , \tag{2.11}$$

which means that any process that maximizes the error frequency around 0 will increase the Correntropy value.

5. Correntropy corresponds to the probability of X=Y. Assuming the joint pdf of the two random variables, X and Y, denoted as $f_{x,y}$, we have

$$f_\varepsilon(0) = \int \ f_{X,Y}(x,x)dx = P(X=Y) \ . \tag{2.12}$$

In other words, the estimate of Correntropy $\hat{V}_\sigma(X,Y)$ with kernel size σ is the integral of estimated $\hat{f}_{X,Y,\sigma}(x,y)$ along the line x=y;

Let's focus now on the CIM, Correntropy Induced Metric. Assuming two vectors, $X=(x_1,x_2,\ldots,x_n)$ and $Y=(y_1,y_2,\ldots,y_n)$, CIM is represented by

$$CIM(X,Y) = (G(0,\sigma^2) - V(X,Y))^{1/2}. \tag{2.13}$$

Assumed as a metric by showing the follow properties:

1. Non-negativity. CIM≥0 ;
2. Identity of indiscernibles. CIM(X,Y)=0 ⇔ X ≡Y;
3. Symmetry. CIM(X,Y)= CIM(Y,X);
4. Triangular inequality. CIM(X,z) ≤ CIM(X,Y)+CIM(Y,Z)

CIM has the following property: because of it non-homogeneous it doesn't induce a norm in the space. However, it behaves as some other metrics depending on the distance to the origin. For instance, in the zero neighborhood the isolines are close to circles, such as the L_2 metric (Euclidian space). far from that neighborhood, the isolines have the square shape, acting like the L_1 metric (sum of coordinates). And finally, in more distant zones, the metric reaches a saturation level and acts like L_0 metric, meaning that it's indifferent to the distance. Such behavior is quite visible in figure 2-4.

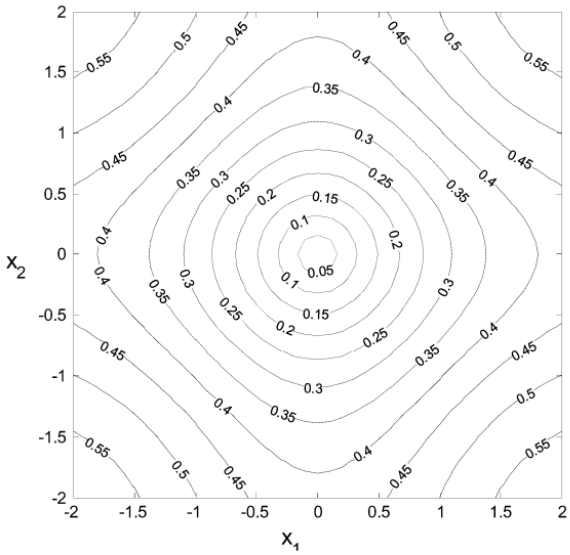

Figure 2-4- Distance contours of CIM(X, 0) in 2-D sample space (kernel size is set to 1). Figure obtained from [9].

It's really interesting to compare the Euclidian metric with the CIM. For that comparison, let's assume two vectors X and Y with N components and each metrics expressions

Euclidian metric

$$d_{L2}(X,Y) = (\sum_{j=1}^{n}(x_i - y_i)^2)^{1/2}. \tag{2.14}$$

CIM metric

$$d_{CIM}(X,Y) = (G(0,\sigma^2) - \frac{1}{n}\sum_{j=1}^{n} G(x_i - y_i, \sigma^2 I))^{1/2} \ . \tag{2.15}$$

In the Euclidean metric, the component that represent a higher influence in its value is the higher difference between X and Y. However, on the CIM, when that difference is too high relating to the parameter σ, it tends to lose its influence on the CIM result. This happens because the Gaussian of a high value (relating to the standard deviation) has a small result. Therefore, the influence of a gross error on the sum of the Gaussians won't be notable.

In this way, for the CIM, the distance between two vectors X and Y will be described by the components with close values, ignoring the components with higher differences. It's important to get clearly that this event depends always on the kernel size, σ. Thus, if the difference between x_i and y_i is already high, it's possible to change the value of the component x_i without affect the value of the mutual CIM. That property is interesting when we don't want to just measure the distance between two vectors, but, when applied to state estimator or other appliances, to exclude outliers without previous treatment.

While the CIM has the advantage to ignore big differences between X and Y, that is, outliers, the L_2 metric has exactly the opposite effect. The L2 result gets affected by that outlier and, as the difference between X and Y increases, the distance value increases also.

2.1.7 - Cauchy-Schwarz inequality

Another interesting measure metric is the Cauchy-Schwarz Inequality. Assuming two probability density functions, $p_1(x)$ and $p_2(x)$, the distance of Cauchy-Schwarz can be given by

$$d_{CS}(X,Y) = -log\frac{\int_{-\infty}^{+\infty} p_1(x)p_2(x)}{\sqrt{\int_{-\infty}^{+\infty} p_1(x)\int_{-\infty}^{+\infty} p_2(x)}}\ . \tag{2.16}$$

This metric has the following properties:

- Symmetric. $d_{CS}(X,Y) = d_{CS}(Y,X)$;
- Is bounded. $0 \leq d_{CS}(X,Y) \leq 1$;
- Only achieves the value 0 when the two functions are exactly the same. Min D_{cs}=0 ⇔ $p_1(x)$=$p_2(x)$.

Applying the expression (2.16) to discrete distributions p_1 and $p_{2,}$ with a set of n points, it turns into

$$D_{CS}(X,Y) = -log\frac{\sum_{i=1}^{n} p_{1i}p_{2i}}{\sqrt{\sum_{i=1}^{n} p_{1i}{}^2 \sum_{i=1}^{n} p_{2i}{}^2}}\ . \tag{2.17}$$

It is possible to observe that in (2.17) the nominator is represented by the calculation for the inner product of the vectors p_{1i} and $p_{2i,}$ and in the denominator are the Euclidean norm of both vectors, as represented in (2.18). Thus, (2.17) can be simplified by (2.19) that is easily convert in (2.20):

$$D_{CS}(X,Y) = -log\frac{\sum_{i=1}^{n} p_{1i}p_{2i}}{\sqrt{\sum_{i=1}^{n} p_{1i}}\sqrt{\sum_{i=1}^{n} p_{12i}}}\ ; \tag{2.18}$$

$$D_{CS}(X,Y) = -log\frac{P_1.P_2}{\|P_1\|\|P_2\|}\ ; \tag{2.19}$$

$$D_{CS}(X,Y) = -log\frac{\|P_1\|\|P_2\|\ Cos\theta_{12}}{\|P_1\|\|P_2\|} = -\log(Cos\theta_{12})\ . \tag{2.20}$$

By the interpretation of (2.20), one must say that Cauchy-Schwarz Inequality depends on the angle between the two vectors (distributions), not depending on the length of the vector.

2.2 -State estimation

This section presents the main aspects related with power system state estimation, discussing fundamental concepts for the developed work. First, the formulation of the state estimation problem is presented, and then, different metrics already applied to its objective function are also exposed. Those different metrics are: WLS to the conventional state estimation process and maximum Correntropy criterion and M-estimators as innovative approaches. The modifications introduced in the state estimation problem caused by the inclusion of PMU is also explored.

2.2.1 - Global concept

The actual electric power system is provided with a Supervisory Control And Data Acquisition system, SCADA, and a EMS, Energy Management System, having a set of monitoring and controlling functionalities. One of these important features is the State estimation [10, 11].

The measured values are usually corrupted with errors and a set of simultaneous measurements can be incompatible due the electrical laws.

State Estimation, SE, aims to process acquired measurements in order to find an estimated valid operational point, with a low error regarding the measured values. In other words, its objective is to find a set of variables (states) that adjusts in the most adequate way to a set of network values (measurements) [12, 13].

The state estimation problem can be formulated mathematically using the following constrained optimization problem:

$$\min \ J(x)\,, \tag{2.21}$$

Subject to

$$c(x) = 0\,, \tag{2.22}$$

$$f(x) \leq 0\ , \tag{2.23}$$

Where:

- x is the vector of state variables;
- J(x) is a scalar function of errors representing the objective function;
- c(x) equality-constraint vector;
- f(x) inequality-constraint vector.

The function J(x) definition is based on a desirable criterion, and it measures the error between the measured values and the estimated values, being that error the parameter desirable to be minimized. The metric mostly used as the conventional state estimation objective function is denominated weighted least square error [14]. The application of new metrics as SE objective function has been theme of discussion. These new metrics includes information theory related concepts, as is was in [1], or m-estimators [15].

The state estimation nonlinear measurement model is formulated using a set of measurement equations. These measurement equations, h(x), relate a set of state variables, x, with a set of measured values, z, and a measurement noise, e, as:

$$z = h(x) + e, \tag{2.24}$$

where:

- h(x) is the non-linear m-dimensional vector function relating states to measurements;
- x is the n-dimensional state vector composed by bus voltage magnitude and phase angles;
- z is the m-dimensional measurement vector. These measurements consist in bus voltage, line power flows and power injection values;
- e is the m-dimensional error vector;

with n<m in order to have an observable system.

The difference between the measured and the estimated value represents the residual error value, represented by

$$r = z - \hat{z} = z\text{-}\ h(x)\,. \tag{2.25}$$

The true value of the measurement error, e, consists in the difference between the measured value and the unknown true value of that component and is calculated by

$$e = z - z^{real} = z - h(x^{true}). \tag{2.26}$$

The true value of the measured component or the state are theoretical and unknown, so it's the error, e, what we know is an estimative to that error, the residual, r.

2.2.2 -- Weighted least square error objective function

As the name of the method suggests, this method's objective is to minimize the square error, assigning weights to each error between the measured and the estimated value. Thus, its objective function is

$$min\ J(x) = [z - h(x)]^T R^{-1}[z - h(x)]\ , \tag{2.27}$$

where R is the m-m-dimension weight matrix, corresponding to the variance of the error and it's obtained by Cov(e) = E[e.e^T]=diag{σ^1,σ^2, …σ^m} [16]. The expected accuracy of the sensor i is used to obtain the standard deviation σ_i of the corresponding measurement i. Therefore, the weight matrix, R, is constructed under the assumption that each sensor measurement has an expected error distribution. More information about the weight matrix can be found in [16].

To solve the objective function represented in (2.27), one needs to find the minimum value of J(x). Therefore, the first-order differential equality has to be observed:

$$g(x) = \frac{dJ(x)}{dx} = -H^T(x)R^{-1}[z - h(x)] = 0\ , \tag{2.28}$$

where H(x) is the Jacobian matrix with m*n dimension.

$$H(x) = \begin{bmatrix} \frac{dh_1(x)}{dx_1} & \cdots & \frac{dh_1(x)}{dx_n} \\ \cdots & \cdots & \cdots \\ \frac{dh_m(x)}{dx_1} & \cdots & \frac{dh_m(x)}{dx_n} \end{bmatrix}. \tag{2.29}$$

Expanding the non-linear function g(x) into its Taylor series around the state vector, x^k, at iteration k, results

$$g(x) = g(x^k) + G(x^k)(x - x^k) + \cdots = 0\ , \tag{2.30}$$

It's possible to get into an iterative solution scheme known as The Gauss-Newton method if the higher order terms are neglected. The iterative function of this method is

$$x^{k+1} = x^k - [G(x^k)]^{-1}.g(x^k)\ \ , \tag{2.31}$$

where:

- k is the iteration order
- x^k is the state vector at iteration k.
- $G(x^k) = \frac{dg(x^k)}{dx} = H^T(x^k).R^{-1}.H(x^k)$
- g$(x^k) = -H^T(x^k).R^{-1}(z - h(x^k))$

The matrix G(x), denominated gain matrix, is a sparse, positive definite and symmetric provided that the system is fully observable [16]. The matrix G(x) is typically not inverted, the state estimate $\hat{x}$ is the obtained by the forward/back substitutions at each iteration k, by

$$[G(x^k)]\Delta x^{k+1} = -H^T(x^k).R^{-1}(z - h(x^k)), \tag{2.32}$$

where $\Delta x^{k+1} = x^{k+1} - x^k$.

The algorithm adopted for WLS state estimation is iterative and terminates when either a maximum number of iterations has been reached or the residual is lower than a defined limit. The iterative process is described in the next points [12, 16].

1. initialize process setting index k=0;
2. Initialize the state vector x^k, a good initialization is important to a quick convergence, typically corresponds to a flat voltage profile, where all bus voltages are assumed to be 1.0 p.u. and all in phase with each other;
3. Calculate the measurement Jacobian, $H(x^k)$;
4. Calculate the gain matrix, $G(x^k)$;
5. Calculate the measurement function , $h(x^k)$;
6. Calculate the correction term, $\Delta x^k = x^k - x^{k-1}$;
7. Test for convergence, $|\Delta x^k| < \varepsilon$?
8. If it didn't converge yet update $x^{k+1} = x^k + \Delta x^k$ and go to step 3. If it already converged stop.

The described procedure assumes that we have some little deviations on measured values and that other information as grid topology, or systems data, are known and well defined, what is not always truth. A set of additional functions are necessary to have a robust state estimator. Some of these functions are:

- topology processing;
- observability analysis;
- bad data processing.

2.2.3 - Correntropy as an alternative estimator

As it was said before, WLS state estimators is the most common criterion and mostly used in the industries. However, new criteria based in theoretic information and M-estimators have been studied, in [1] and [15], respectively.

Correntropy is a proper measure of the similarity between two random variables X and Y, and can be expressed as in [8]

$$V_\sigma(X,Y) = E[G(X-Y,\sigma^2 I)] \ . \tag{2.33}$$

Applying this criterion to the state estimation process, one must see X as the measured value and Y as de estimated value, having the residuals R= X-Y. The number of measurements is limited and discreet, thus, also it is the number of estimation. In this way, Correntropy has to be calculated based on those discreet points, by

$$\hat{V}(X,Y) = \frac{1}{n}\sum_{j=1}^{n} G(x_i - y_i, \sigma^2 I) = \frac{1}{n}\sum_{j=1}^{n} G(\varepsilon_i, \sigma^2 I). \tag{2.34}$$

It's remarkable an affinity between the Correntropy and the Parzen window method. Actually, by adjusting the parameter σ, the window that defines the visible spectrum of errors is adjusted. Therefore, if ε_i is much bigger than σ, that error is seen as an outlier and the component *i* is ignored. The convenient selection of the PW size allows to naturally identify and ignore components whose measurement contains a gross error.

This is the big difference between the Correntropy and the WLS. While on the WLS a gross error pulls the estimations of all the other components in the wrong direction, in the Correntropy the gross error is ignored, allowing all the other elements to conduct to the right estimation [1].

The state estimation objective function using the maximum Correntropy criterion is as below, and leads to a minimum error between the estimations and the measured values.

$$MCC(\varepsilon) = max\ \frac{1}{n}\sum_{j=1}^{n} G(\varepsilon_i, \sigma^2 I), \tag{2.35}$$

where n is the number of measured components, ε_i is the error between the measurement and the estimation for the component i and $\sigma^2 I$ is the standard deviation matrix .It leads to a minimum error between the estimate and the measured values, ignoring possible outliers without any requirement for pretreatment of the measurement data.

2.2.4 -M-estimator as an alternative estimator.

In [15], Such as in maximum Correntropy criterion [1], was explored also an exponential criterion, but from the perspective of M-estimators.

M-estimator is a robust technique used in state estimation. It tries to decrease the effect of gross errors by replacing the square error (used in the least square error), by another function of errors. In the m-estimator perspective, the objective function formulation may be translated by:

$$min \sum_{i}^{n} \rho(r_i) \tag{2.36}$$

Where ρ is a function of the residual error. This function is symmetric, positive defined and has only one minimum, defined at zero. Its formulation is provided to be less increasing than the square function [17].

As the weighted least square [16] or the reweighted least square [18], this method can be implemented by an iterative process, where weights are assigned to each residual error. The weight function is defined by:

$$\omega(r_i) = \frac{\Psi(r_i)}{r_i}. \tag{2.37}$$

The influence function $\Psi(x)$, present in the previous expression, can be translated by

$$\Psi(r_i) = \frac{d\rho(r_i)}{r_i}, \tag{2.38}$$

and it measures the influence of a component's error on the estimation.

The non-robustness of least-square estimator can be explained by its influence function. In the WLS case the influence function is $\Psi(r_i) = r_i$, Which means that the influence of an component's error on the estimation increases linearly with its error. Only when the influence of any single error is insufficient to yield any significant offset, the estimator can be assumed as robust. [19].

To solve (2.36) one must find its minimum. As $\rho(r_i)$ is differentiable and has only one minimum, the differential process must provide the solution.

$$\frac{d\rho(r_i)}{dx} = \frac{d\rho(r_i)}{d(r_i)}\frac{d(r_i)}{dx} = 0, \tag{2.39}$$

applying (2.38) to (2.39), it simplifies as

$$\frac{d\rho(r_i)}{dx} = \Psi(r_i)\frac{d(r_i)}{dx} = 0, \tag{2.40}$$

and applying (2.37) to (2.40), finally we have

$$\frac{d\rho(r_i)}{dx} = \omega(r_i).r_i.\frac{d(r_i)}{dx} = 0. \tag{2.41}$$

As it is exposed in [17], the equation system (2.41) corresponds to the same equation system of the following iterative reweighted least square:

$$\min \sum_i w(r_i^{k-1})r_i^2, \tag{2.42}$$

where k is the iteration index. The same equation system can be observed in the WLS

$$\min \sum_i W_i r_i^2. \tag{2.43}$$

This demonstration proves that the m-estimator problem can easily be solved the same way as the iterative WLS is.

A M-estimator requires to present some relevant constraints. Through [17], these requirements are:

- Its influence matrix must be bounded;
- The error function $\rho(r_i)$ must be convex in the covered space, having a unique minimum ;
- Whenever the objective function second derivate $\frac{d\rho^2(r_i)}{dx^2}$ is singular, its gradient must be different than zero. This requirement aims at limiting the search space.

In [15] the author uses an objective function mathematically identical to the maximum Correntropy criterion. The difference is the weighted exponential and the interpretation of M-estimators. The objective function presented is:

$$\max J(x) = \sum_{i=1}^{m} w_i \exp\left(-\frac{\left(z_i - h_i(x)\right)^2}{2\sigma^2}\right) = \sum_{i=1}^{m} w_i \exp(-r_{si}^2), \tag{2.44}$$

where σ is a constant, w_i is the weight at i^{th} sample and $r_{si} = \frac{z_i - h_i(x)}{\sqrt{2}\sigma}$.

There are two important particularities of this M-estimator (2.44). First, it is continuously differentiable, allowing the application of conventional optimization methods as Gauss-Newton. Second, it has the ability to suppress bad data.

The $\exp(-r_{si}^2)$ from equation (2.44) can be seen in the following image 2-5. Here is notorious that for a small error, $r_{si} \simeq 0$, the influence of the component *i* in the final results will be notable. For a gross error, r_{si} is relatively large, thus $\exp(r_{si}) \simeq 0$. This means that the component seen as an outlier will not influence the estimation results, ensuring the robustness of the method. This peculiarity is also noticed when using the maximum Correntropy criterion.

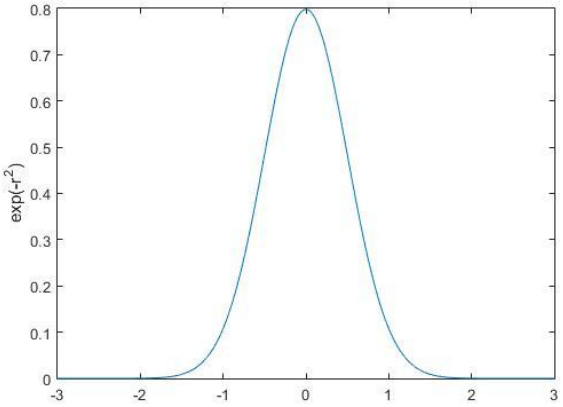

Figure 2-5 - Exponential square objective function for one residual.

In [15] is still possible to observe the SE based in an special m-estimator being solved by the Gauss-Newton method previously explained in section 2.2.2 -, as well as some experimental results.

2.2.5 -Measurement equations and influence of PMU on the Jacobian matrix

The conventional measurements correspond to power injections, branches power flows and bus voltage magnitudes. If we assume the DC model the state variables are the bus phases. As measurements, only real power and voltage phases are considered. The measurement function for power injections is

$$P_{iinj} = [B][\theta]\,, \tag{2.45}$$

where [B] is the admittance matrix for the DC simplification model (where each component corresponds to $-1/x_{ij}$) and [Θ] is the bus voltages phases vector. For line power flow the expression is

$$P_{ij} = \frac{\theta_i - \theta_j}{x_{ij}}\,, \tag{2.46}$$

where Θ_k is the voltage phase at bus k and x_{km} is the reactance of the line k-m.

The AC model measurements functions, assuming only conventional sensors, are following presented in equations (2.47) to (2.50). For power injections the expressions for active and reactive power are, respectively,

$$P_i = V_i \sum_{j\epsilon Nj} V_j(G_{ij}cos\theta_{ij} + B_{ij}sin\theta_{ij}) \text{ and} \tag{2.47}$$
$$Q_i = V_i \sum_{j\epsilon Nj} V_j(G_{ij}sin\theta_{ij} - B_{ij}cos\theta_{ij}), \tag{2.48}$$

Where V_i is the voltage magnitude at bus *I*, G_{ij} and B_{ij} are respectively, the conductance and susceptance for the line ij and $\Theta_{ij}=\Theta_i - \Theta_j$. The lines power flows the expressions are

$$P_{ij} = V_{ij}^2(g_{si} + g_{ij}) - V_iV_j(g_{ij}cos\theta_{ij} + b_{ij}sin\theta_{ij}) \text{ and} \tag{2.49}$$
$$Q_{ij} = -V_{ij}^2(b_{si} + b_{ij}) - V_iV_j(g_{ij}sin\theta_{ij} - b_{ij}cos\theta_{ij}). \tag{2.50}$$

PMU introduces new measurements to the system. Actually, it has the ability not only to measure bus voltage phases, but also the current phasor in the lines connected to the bus where the PU is installed. Thus, to the conventional measurement vector will be added voltage

phase angle and current phasor. The approach used in [12,13] considers the current injected in the buses with PMU. Obviously it corresponds to the summation of the line currents adjacent to that bus. Therefore, the Jacobian matrix will have four more rows, and will be as follows.

$$H(x) = \begin{bmatrix} \frac{dP_{ij}}{d\theta} & \frac{dP_{ij}}{dV} \\ \frac{dP_{ij}}{d\theta} & \frac{dP_{i}}{dV} \\ \frac{dQ_{ij}}{d\theta} & \frac{dQ_{ij}}{dV} \\ \frac{dQ_{i}}{d\theta} & \frac{dQ_{i}}{dV} \\ \frac{dV_{i}}{d\theta} & \frac{dV_{i}}{dV} \\ \frac{d\theta_{iPMU}}{d\theta} & \frac{d\theta_{iPMU}}{dV} \\ \frac{dV_{iPMU}}{d\theta} & \frac{dV_{iPMU}}{dV} \\ \frac{dI_{iReal}}{d\theta} & \frac{dI_{iReal}}{dV} \\ \frac{dI_{iImag}}{d\theta} & \frac{dI_{iImag}}{dV} \end{bmatrix}. \quad (2.51)$$

The expression for the bus injected currents is

$$Ii = \sum_{j \epsilon Nj} y_{ij} V_j, \quad (2.52)$$

and can be represented by the rectangular components calculating the real and imaginary part as

$$Ii(real) = \sum_{j \epsilon Nj} (V_j (G_{ij} \cos(\theta_j) + B_{ij} sin(-\theta_j)) \text{ and} \quad (2.53)$$

$$Ii\ (imaginary) = \sum_{j \epsilon Nj} (V_j (G_{ij} \sin(-\theta_j) - B_{ij} \cos(-\theta_j)). \quad (2.54)$$

The inclusion of PMU in the measurement system introduces conflicts between quantities expressed in rectangular coordinates and polar coordinates. In a conventional estimator the system states are expressed in polar coordinates, which means that the injected currents have to be expressed by nonlinear functions of magnitude and phase angle Voltage at buses. Another option is to assumed that PMU measures the magnitude and the phase angle of line currents (which are still nonlinear functions of the system state). Although it's possible to convert rectangular coordinates into polar, the problem lays on the characterization of the covariance matrix of measurement errors [12,13].

It's important to mention that the equations presented in this subchapter are the functions that relates the state variables with the estimations of each one of the measured value

2.3 - Sensory fusion

Data Fusion with Multiple Sensors can be seen as "the theory, techniques and tools which are used for combining sensor data, or data derived from sensor data, into a common representational format" [20]. It is commonly compared to the way humans combine different senses as vision, smell, touch and taste to interact with the environment around.

Sensory fusion allows the system to be more tolerant to measurements faults and to own a capability to provide new information in a way that would not be possible with only one sensor. In order to overcome the existing limitations when using only one sensor ,the purpose of fusion methods is to use all the diverse information acquired by multiple sensors [21]. The fusion applications have been focused on increasing the quality of produced values through the processing means. Particularly in instrumentation, sensory fusion is used to decrease the measurements uncertainty. Thus it increases the reliability of the obtained values through combining of individual measured data [22].

Most of the times sensory fusion is supported by a statistical approach. The main advantage of these approaches is: explicit probabilistic models are applied to characterize the relationships between sensors and sources of information, taking into account the subjacent uncertainties [21].

Currently, in electric power systems there is no specific algorithm to fusion different types of sensors. What is mostly used is a WLS estimator. This estimator has a weight matrix formulated according to the covariance of each expected sensor error [2].

2.3.1 - Synergy

Synergy is a process that improves the quality of the information output. Synergy does not imply the use of multiple sensors. Truly speaking, it may be provided on a temporal sequence obtained by a unique sensor. However, the use of multiple sensor may increase the effect of the synergy in different ways.

Sensor fusion is used in the most different areas with the purpose to get more reliable information. For instance, in multi-modal biometric systems, when using only one biometric trait for recognition, error rates reaches high values. These big errors are caused by the lack of completeness or universality in biometric traits, as it is the case of fingerprints. When analyzing people with hand related disabilities, cuts or burns on the fingertips or people with very dry hands is not possible to get a good quality fingerprint. Thus, fingerprints are not truly universal. The way to solve this problem is to use multi-modal biometric sensor system by fusing the data obtained by the different traits [23].

Trough [20] the multi-sensor data fusion may get better performance of the system in four different ways:

- Representation. The information obtained after the fusion process has an abstract level lower that each input of the data set.
- Certainty. If we assume p(V) as the *a priori* probability of a data vector V before fusion, we expect a gain in certainty represented by the growth in p(V). Therefore, it's expected $p(V_f)>p(V)$, where V_f is a vector of data after fusion.
- Accuracy. Usually, the gain in accuracy and certainty are correlated. The fusion process tries to reduce the errors and noise present in the sources. In this way, the standard deviation on the data after the fusion process is smaller than the standard deviation on the original data.
- Completeness. Analyzing more information about an environment allows a more complete perspective about it. If that new information is redundant and concordant, we will get also a gain in accuracy.

2.3.2 - Multi sensor data fusion strategies

In [20] it considers three different points of view across data fusion suggested by Boudjemaa and Forbes, Durrant-Whyte and Dasarathy [24-26] .

Boudjemaa and Forbes distinguishes the type of data fusion according to which part of the system is fused. More specifically, if it is fused across sensors, attributes, domains or time. In fusion **across sensors,** it's assumed that all the sensors measure the same property, while in fusion **across attributes,** sensors measure different properties. However, all the properties are related to the same experimental situation. For example, to calculate the refractive index it's necessary to measure: air temperature, pressure and humidity. In fusion **across domains,** the sensors measure the same attribute across different domains or ranges. Lastly, in fusion **across time,** it's used historical information that will be fused with actual data.

Durrant-Whyte defines the data fusion system according with the sensor configuration. The classification of the system is either complementary, competitive or cooperative. It is considered **complementary** if the sensors do not depend directly on each other, but them information can be combined to get a better image of the system. Incompleteness problem can be improved by complementary sensors. In **competitive** configuration, a certain number of sensors measures the same property, aiming to reduce the effect of uncertain and erroneous measurements. Finally, **cooperative** sensor configuration uses the information provided by at least two independent sensors. The combination of that data must generate information that would not be available from the single sensors.

Dasarathy has a perspective across the input/output characteristics, having five different classifications:

- DaI-DaO- Data Input/Data Output. In this type of data fusion, the input data is filtered;
- Dai-FeO- Data Input/Feature Output. Input data generates output Features;
- Fei-FeO- Feature Input/Feature Output. Input features are decreased in number or new features are generated by input features;
- Fei-DeO- Feature Input/Decision Output. Input features are fused to give out-put decision;
- Dei-DeO- Decision Input/Decision Output. A final out-put decision is given by the fusion of multiple input decisions.

2.3.3 - Catastrophic Fusion

Multi-sensor fusion system helps to improve the results in comparison with a single sensor system, but not always. Sometimes the performance of a multi-sensor fusion system is lower than the performance of individual sensors. This phenomenon called catastrophic fusion should be always avoided [20].

In general each sensor S_m,with mϵ {1,2,…,M}, is designed to be in service under specific conditions C_m. If C_F is the correct operation of all the sensors in the system F, it can be calculated by

$$C_F = C_1 \wedge C_2 \wedge \dots C_m. \tag{2.55}$$

However, sometimes it can happen that the system F is used under inconsistent conditions. Denoting the inconsistent condition as C_{m^*}, The signal from the sensor S_{m^*} may corrupt the fused output with catastrophic results. This is exactly what happen with the WLS estimator. When one of the measurements is affected with a gross error, the condition of the subjacent error distribution being a Gaussian is infringed. Therefore, all the fusion will be strongly affected by that component. To prevent this phenomenon, multi-sensor fusion systems must incorporate secondary classifiers which monitor the performance of each sensor S_m.

2.3.4 - Sensor data fusion process

Sensory fusion has different process according to the area where it is inserted. In [27] there is a report about aircraft sensor data fusion. It describes the sensory fusion process in that field. The sensory fusion process is divided in two states. The first one is the tracking stage. In this stage all the sensors measurements referring to a platform (fighter aircraft engaged in an air mission) are brought together over time. Thus, it gives the most complete and accurate estimate of the platform's position, motion and status that the measurements allow. The desirable would be to gather all the sensor measurements into a single tracking process in a timely and reliable way. Thus, a single track database would be produced and the sensory fusion would be complete at this stage. However, there are a lot of constraints on the systems and this doesn't happen. In the end of this first stage several tracking process have produced them own database, each with a different perspective or image of the observed system, these different images must be combined in the next stage.

The second stage, Track-to-Track Fusion, aims to fusion all these track databases into only one fused database. To fusion all the databases it's necessary to perform the three next tasks:

- Align the tracks to the same spatial axis set and time;
- Deduce the number of targets that originated the acquired data and the platform from which each track arose;
- Form joint tracks and joint identity statements for targets reported by more than one source.

In [28] is proposed a distribution state estimator based on the autoencoders for three-phase state estimator. The proposed method presents 3 main steps:

- Constructing the historical data set;
- Training the autoencoder (offline procedure);
- Building the distribution state estimator algorithm.

Briefly speaking, only after define a set of historical data and train the autoencoder(as a neural network), a state estimation with the actual measurements is provided. In a certain way, can be said that there is a fusion across time, since historical data is influencing the estimation process for the actual measurements. However, all the actual measurements are gathered together, and the state estimation is provided without any kind of consideration or distinction about the sensor that provided the measurement.

In other works, as [12-13], even existing distinct sensor types, the measured data is collected all together. The only distinction is through the weight matrix from the WLS estimator. There is no special attention or distinguish about the sensory system that provided the measure, and other parameters as sensory system's reliability.

Chapter 3

Comparison between different state estimation optimization functions

The mostly used criterion as objective function of power system state estimation is the weighted least square error [14]. The application of the WLS minimizes the variance of probability density function (pdf) of the residual error distribution. A good performance of the state estimator under the WLS criterion requires a Gaussian error distribution, what's not always true. Sometimes the measurements contain gross errors. Such bad data affect the WLS solution defiling the estimative of components with healthy measurements. Once WLS cannot recognize and discard these large errors, it cannot be considered a robust estimator.

Therefore, in this chapter, the traditional WLS will be compared with information theory related concepts. This chapter is dedicated to the application of those concepts as state estimation (SE) optimization function. Basically, the idea is to find a criterion able to maximize the information extracted from the available measurements. In other words, it aims at maximizing the information on the estimations by minimizing the informational content of the error distribution. If we achieve an error distribution with no information, the error pdf corresponds to a Dirac function centered in zero, meaning that there is only one solution for the state variables, which explain all measured values perfectly. That's what would happen with a minimum entropy.

In this way, a criterion based on theoretic information minimizes the informational error distribution, while the WLS minimizes the variance of the error distribution. The advantage of information theory related criteria is that, unlike the WLS that only consider the second order moments, it uses all the moments of the error distribution, providing more reliable results, as it was already explain in chapter 2.1.6 -.

Correntropy, one of the alternative criteria, shows really interesting properties that can be easily applied in state estimation. MCC has the particularity of changing the perception of similarity according with the distance between the measured and estimated value. That is, depending on that distance, it behaves as different norms, namely: Euclidian-L2, absolute distance-L1 and then a saturation can be observed approaching L0. Thus, MCC gives more attention to close values and, then, gradually ignore the components as the residual increases. Therefore, MCC is able to ignore measurements corrupted with large errors. In this way, Correntropy is proposed as robust method to measure the similarity between measured and

estimated values. Actually, this criterion was already tested in power system state estimation, in [1], promising really interesting results as a robust state estimator.

In this chapter others criteria are discussed beyond the Correntropy, such as Renyi's quadratic entropy and Cauchy-Schwarz distance.

In order to evaluate the characteristics of the different metrics, two different scenarios are considered:

- Measurements without gross errors;
- Measurements with one gross error.

For the scenario with gross error, two different cases were compared. A gross error in injected power in bus 3 and bus 9, respectively.

3.1 - Application of the alternative criteria in state estimation problem

The first experiment in this thesis consists in comparing some different criteria as state estimation objective function. The evaluated criteria in this chapter are: Correntropy Induced Metric (also used as maximum Correntropy criterion) (3.1), Renyi's quadratic entropy (3.4), Cauchy-Schwarz distance (3.10) and, as a comparison, the traditional least square error.

The main objective of the state estimation is to minimize the distance between measured values and its estimations, being possible to measure that distance by different metrics.

Two vectors are considered in the following expressions: the measurement vector Z with n components z_i, with iϵ{1,2….n}, and the estimation vector $\hat{Z}$ with the n components $\hat{z}_i$, where iϵ{1,2….n}. The difference between this two vectors is the residual vector, that is, the deviation of the estimation relating to the measured value, $r_i = z_i - \hat{z}_i$.

Using the CIM, the objective is to minimize the distance between the estimation and the measured value, given by

$$d_{CIM}(Z,\hat{Z}) = (G(0,\sigma^2) - V(Z,\hat{Z})^{1/2} = (G(0,\sigma^2) - \frac{1}{n}\sum_{i=1}^{n} G(z_i - \hat{z}_i, \sigma^2 I))^{1/2} \ . \quad (3.1)$$

However, it is possible to see that the expression (3.1) has a constant inside. The constant remains always the same, what matters is the variable term. Thus, the term $G(0,\sigma^2)$ can totally be ignored, just as the fraction term 1/n. Is still possible to make another simplification: To minimize the square root of a positive defined function is exactly the same as minimize that function. Thus, (3.1) can be simplified by

$$d_{CIM'}(Z,\hat{Z}) = -\sum_{j=1}^{n} G(z - \hat{z}_i, \sigma^2 I) = -\sum_{j=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z_i - \hat{z}_i)^2}{2\sigma^2}} \ . \quad (3.2)$$

Now is possible to convert the minimization problem into a maximization problem, just by convert the negative function into a positive. It visible that the term $\frac{1}{\sigma\sqrt{2\pi}}$ is the same in all the components inside the summation and it can also be ignored, leading to

$$d_{CIM}(Z,\hat{Z}) = \sum_{j=1}^{n} e^{-\frac{(z_i - \hat{z}_i)^2}{2\sigma^2}} \ . \quad (3.3)$$

In this way, applying Correntropy to the state estimation problem, the objective function is the maximization of (3.3). From now on, the Correntropy used as the SE criterion will be

denoted MCC, Maximum Correntropy Criterion. In chapter 5, Correntropy will be used to measure the distance between two points, in that case, will be referred as CIM. The reason for this choice was to avoid confusion between the two simultaneous employments of this metric.

Now, applying the Renyi's quadratic entropy, the objective function is the minimization of the error distribution information. Once the true error is unknown, the objective is to minimize the residual error information by

$$H_2 = -log \int_{-\infty}^{\infty} f(r)^2 dr \,, \tag{3.4}$$

or the maximization of the symmetric function of (3.4), where f(r) represents the pdf of the errors.

In effect, the f(r) true distribution is unknown, what is available is a set of discrete residual errors. The residual error distribution, f(r), can be estimated by the PW method from a discreet sample of error values. It consists in applying in each residual a Gaussian kernel as

$$\hat{f}(r) = \frac{1}{n}\sum_{i=1}^{n} G(r - r_i, \sigma^2 I) \,, \tag{3.5}$$

where n is the number of samples and σ is the size of the Parzen windows, given by the standard deviation of the adopted Gaussian Kernel. Besides the estimation of the error distribution, the application of the PW method to the RQE concept also intends to introduce a parameter able to manage the expected error distribution. Thus, the proper definition of the parameter σ must allow the identification and natural elimination of large errors. Applying the estimate $\hat{f}(r)$ to (3.4) results

$$H_2 = -log \int_{-\infty}^{\infty} \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} G(r - r_i, \sigma^2 I)\, G(r - r_j, \sigma^2 I) \,, \tag{3.6}$$

that is the same as

$$H_2 = -log \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \int_{-\infty}^{\infty} G(r - r_i, \sigma^2 I) G(r - r_j, \sigma^2 I) \,. \tag{3.7}$$

In (3.7) we came across the integral of the product of Gaussians, which may be assimilated to the convolution of Gaussians, with a well-known result: The Gaussian of the difference [7]:

$$H_2 = -log \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} G(r_i - r_j, \sigma^2 I), \tag{3.8}$$

$$H_2 = -log \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(r_i - r_j)^2}{2\sigma^2}}. \tag{3.9}$$

In (3.9) we have the final form of RQE applied to state estimation problem. The constants outside the summation don't interfere in the optimization process and the maximization of a logarithm is the same as maximize the term inside the logarithm. Therefore, it's important to mention that minimize that expression is the same as maximize only the double sum of the exponential term. Thus, the objective function is the one presented in table 3-1.

As explained in chapter 2.1.7 - the Cauchy-Schwarz inequality can be calculated by

$$D_{CS}(Z, \hat{Z}) = -log \frac{\sum_{i=1}^{n} z_i \hat{z}_i}{\sqrt{\sum_{i=1}^{n} {z_i}^2 \sum_{i=1}^{n} {\hat{z}_i}^2}} \tag{3.10}$$

and can be interpreted as a measure the collinearity between two vectors, by the simplification

$$D_{CS}(Z,\hat{Z}) = -\log(Cos\theta_{z\hat{z}}), \tag{3.11}$$

depending only on the angle between the two compared distributions. Once again, the exclusion of the signal minus from the expression (3.10) turns the minimization problem into a maximization problem.

The distributions we want to compare, or to approach, are the measurement distribution and its estimation, Z and $\hat{Z}$, respectively.

Summing up, the set of different objective functions that are going to be applied to the SE process can be consulted in table 3-1.

Table 3-1- Different objective functions.

Weighted Least Square Error	$\mathbf{min}\,[\boldsymbol{Z}-\widehat{\boldsymbol{Z}}]^{\boldsymbol{T}}\boldsymbol{R}^{-1}[\boldsymbol{Z}-\widehat{\boldsymbol{Z}}]$
Correntropy (MCC)	$max\sum_{j=1}^{n} e^{-\frac{(z_i-\hat{z}_i)^2}{2\sigma^2}}$
Renyi's Quadratic Entropy (RQE)	$max\sum_{i=1}^{n}\sum_{j=1}^{n} e^{-\frac{(r_i-r_j)^2}{2\sigma^2}}$
Cauchy-Schwarz inequality	$\max\frac{\sum_{i=1}^{n} z_i\hat{z}_i}{\sqrt{\sum_{i=1}^{n} {z_i}^2\sum_{i=1}^{n}{\hat{z}_i}^2}}$

It's important to mention that, the construction of the weight matrix (from WLS) requires information about the error covariance, for the multiple components. For a matter of simplification, in this work was assumed that all the errors had the same variance. Thus, the matrix R can be ignored. Actually this simplification will not interfere with the following studies. Assuming this simplification, the tested criterion was the least square error, or minimal mean squared error (MSE). Because in the also in MCC and RQE could be integrated a weight matrix, and at the moment it is not considered, the comparison with MSE approach is fairer than WLS. From now on, the minimal MSE will be used to detonate the traditional SE optimization function.

3.2 -Network and measurements for the DC model state estimator

The different criteria were tested in the network presented in the annexes (A). This adaptation is based in a typical European medium voltage network adapted from [29]. In figure 3-1 is possible to see the network topology and, in table 3-2, the measurement set for the DC model. For more information, consult Annex A.

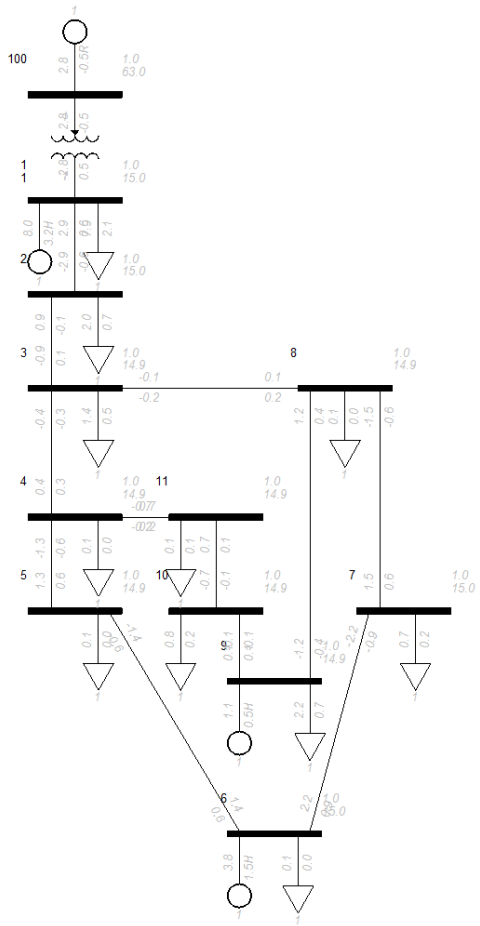

Figure 3-1 - Network used to compare the different criteria.

Table 3-2 - Measurements used to compare the different criteria .

	Bus/line	Power Flow result	Measurement affected with noise
P_{inj} (p.u.)	100	0.111326	0.111469
	1	0.004130	0.004325
	2	-0.080749	-0.080625
	3	-0.056928	-0.056763
	4	-0.004777	-0.004955
	5	-0.004907	-0.005229
	6	0.144789	0.144945
	7	-0.029052	-0.028692
	8	-0.005973	-0.006022
	9	-0.042991	-0.042779
	10	-0.031513	-0.031764
	11	-0.003354	-0.003434
θ (rad)	2	-0.013365	-0.013383
	5	-0.014911	-0.014913
	8	-0.015388	-0.015403
P_{ij} (p.u.)	3-4	-0.017071	-0.016890
	6-7	0.088141	0.087880

3.3 - State Estimation without gross errors

The state estimation problem was solved with the assistance of EPSO. The state variables were the voltage phases. The tested objective functions were the ones previously presented in the table 3-1.

Table 3-3- State estimations provided by each criterion.

Component		Z	$\hat{Z}_{MSE}$	$\hat{Z}_{MCC}$	$\hat{Z}_{RQE}$	$\hat{Z}_{CS}$
P_{inj} (p.u.)	100	0.111469	0.111375	0.111375	0.111376	0.000341
	1	0.004325	0.004241	0.004241	0.004240	0.000013
	2	-0.080625	-0.080704	-0.080703	-0.080706	-0.000247
	3	-0.056763	-0.056836	-0.056836	-0.056840	-0.000174
	4	-0.004955	-0.004976	-0.004977	-0.004975	-0.000015
	5	-0.005229	-0.005294	-0.005294	-0.005294	-0.000016
	6	0.144945	0.144769	0.144769	0.144765	0.000443
	7	-0.028692	-0.028560	-0.028559	-0.028556	-0.000087
	8	-0.006022	-0.006016	-0.006016	-0.006016	-0.000018
	9	-0.042779	-0.042777	-0.042777	-0.042776	-0.000131
	10	-0.031764	-0.031773	-0.031774	-0.031772	-0.000097
	11	-0.003434	-0.003448	-0.003449	-0.003446	-0.000011
ϴ (rad)	2	-0.013383	-0.013406	-0.013406	-0.013406	-0.000041
	5	-0.014913	-0.014972	-0.014972	-0.014972	-0.000046
	8	-0.015403	-0.015447	-0.015447	-0.015447	-0.000047
P_{ij} (p.u.)	3-4	-0.016890	-0.016803	-0.016802	-0.016807	-0.000051
	6-7	0.087880	0.088203	0.088203	0.088199	0.000270

Table 3-4- Residual errors obtained from each criterion.

Component		R_{MSE}	R_{MCC}	R_{RQE}	R_{CS}
P_{inj} (p.u.)	100	-0.000094	-0.000094	-0.000093	-0.111128
	1	-0.000084	-0.000083	-0.000085	-0.004312
	2	-0.000079	-0.000079	-0.000081	0.080378
	3	-0.000073	-0.000073	-0.000077	0.056589
	4	-0.000021	-0.000021	-0.000019	0.004940
	5	-0.000065	-0.000065	-0.000065	0.005213
	6	-0.000176	-0.000176	-0.000180	-0.144502
	7	0.000132	0.000133	0.000136	0.028604
	8	0.000006	0.000006	0.000006	0.006004
	9	0.000002	0.000002	0.000003	0.042648
	10	-0.000010	-0.000011	-0.000008	0.031666
	11	-0.000014	-0.000015	-0.000012	0.003423
ϴ (rad)	2	-0.000023	-0.000023	-0.000024	0.013342
	5	-0.000059	-0.000059	-0.000059	0.014867
	8	-0.000044	-0.000044	-0.000044	0.015356
P_{ij} (p.u.)	3-4	0.000087	0.000088	0.000083	0.016839
	6-7	0.000323	0.000323	0.000319	-0.087610

The estimated values obtained using the different criteria can be consulted in table 3-3 and its residual error in table 3-4. For the MCC and RQE criteria it was utilized an PW size 5.

For a better illustration of the results, in the following charts is possible to see the residuals distribution obtained though the PW method. That distribution was obtained by placing a Gaussian kernel, with a standard deviation of 0.01, in each of the residual values. Summing all those terms and divide by the number of summed terms was obtained the residual error pdf's presented in figures 3-2 and 3-3.

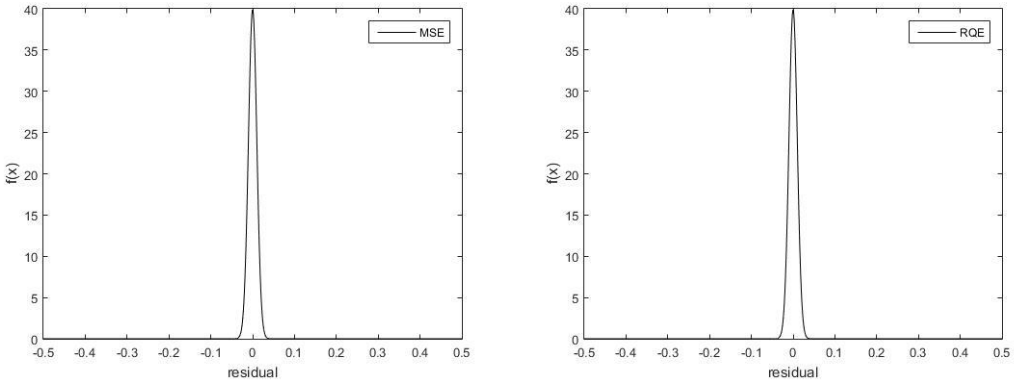

Figure 3-2 – Residual errors distribution for MSE, on the left , and RQE , on the right.

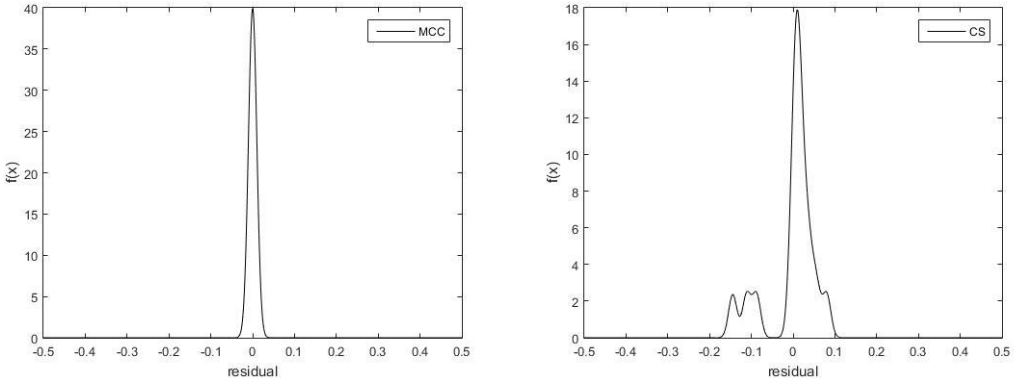

Figure 3-3 – Residual errors distribution for MCC, on the left , and CS , on the right.

3.3.1 - Results analysis

The first big conclusion is about the Cauchy-Schwarz distance. The error of the estimation provided by this metric are at the same order as the measured values, revealing to be much higher than the errors obtained with the other tested metrics. It is also possible to observe that it's distribution, figure 3-3, doesn't follow a Gaussian distribution as expected and as the other metrics follow. In the Cauchy-Schwarz is possible to see peaks and a distorted distribution caused by the bad estimation.

In fact, as explained before, this metric measures the distance between two distributions according with them collinearity. Thus, the application of CS metric in a state estimation problem leads to a trouble situation: it will try to find an estimation vector collinear with the measurement vector, which can lead to a solution far from the optimal one. That is what happened in the previous example. The found estimation vector is collinear with the measurement vector; However, the "optimal" point found has a high error comparing to the

residual errors obtained by the other metrics. Despite the other metrics provided estimation vectors not collinear with measurement vector, the estimations were closer to the measured values. For now on, the Cauchy-Schwarz metric will be excluded once it was proved that collinearity is not a good criterion to the SE problem.

When the measurements do not contain gross errors, Correntropy and Renyi's quadratic entropy criterions conduct to an optimal estimation, near to the measurement vector. Also, a similarity between Correntropy and MSE results is notable. In fact, when there are no gross errors the Correntropy metric presents the same result as (or really close to) MSE, once MSE minimizes $(z_i-\hat{z}_i)^2$ and, in a certain way the Correntropy minimizes also that component. Maximize the exponential of the negative square error (divided by a constant) is the same as minimizing the square error. The big difference is that the Correntropy has a division term affecting the square error, and to make this effect even more notable, also an exponential. These additional terms in Correntropy have effect in situations with gross errors, as we will see in the next section.

3.4 - State estimation with gross errors

One problem with the MSE state estimator is its reaction to bad data. When facing outliers, the component that contains a big error infects also the other components estimation, deallocating all the estimations in its direction. For that reason, in order to avoid bad data and to perform a proper estimation, the Least Square Error state estimator requires a pretreatment of the acquired data.

Therefore, it would be interested to evaluate the other metrics performance under the presence of large errors. An important parameter that requires special attention is the Parzen Window size for the identification of outliers. The explore the behavior of the different criteria, it was used the measurements presented in 3.2 with a gross error in Buss 9 and other case with a gross error in bus 3.

The reason to compare two different cases with large errors in two different buses is to assess its influence in adjacent line power flows estimation. In the first study case a gross error was introduced in a bus without measurements power flow in adjacent lines, bus 9, and, in the second study case, a large error was introduced in bus 3, where an adjacent line power flow is available.

From the analysis of tables 3-5 and 3-6, it has become evident that, facing a gross error, the MSE results are corrupted and all the estimations are moving towards that error. The same happens with Correntropy and Renyi's quadratic Entropy when the Parzen Window size is high comparing to the error magnitude (in Correntropy) or to error distribution variance (in Renyi's entropy), case σ=5.

In Correntropy, when σ=5, the Parzen Window size is too large and the error introduced in 9 "fits" inside it. In other words, because the introduced error is too small comparing to the PW, it is not seen as a gross error but as an available measurement, as all the others. When the Parzen Window size decreases, σ=0.1, the error introduced in bus 9 is seen as an outlier. In fact, the value of the error is intensified dividing by 0.1, then applying the negative exponential, results a really small component compared to the other components, calculated from smaller errors. Thus, because the influence of the measured value P_{inj_9} doesn't have an influence to the summation of all the terms of the Correntropy expression, that component is ignored. This is the reason why, for σ=0.1, all the errors in the other components are really

small. The only big error (between the estimation and the measured value) is the one in injected power in bus 9. In this way, that measurement was ignored and the Correntropy provided a proper estimation ignoring the outlier.

Table 3-5- State estimations provided by each criteria, with different PW sizes. Gross error in P_{inj9}.

Component		Z	$\hat{Z}_{MSE}$	$\hat{Z}_{MCC}$ (σ =5)	$\hat{Z}_{RQE}$ (σ =5)	$\hat{Z}_{MCC}$ (σ =0.1)	$\hat{Z}_{RQE}$ (σ =0.1)
P_{inj}	100	0.111469	0.044074	0.044051	0.062641	0.111374	0.111352
(p.u.)	1	0.004325	-0.112388	-0.112369	-0.124765	0.004239	0.004198
	2	-0.080625	-0.221311	-0.221282	-0.248725	-0.080705	-0.080748
	3	-0.056763	-0.234357	-0.234333	-0.277623	-0.056837	-0.056889
	4	-0.004955	-0.190753	-0.190738	-0.171085	-0.004978	-0.005016
	5	-0.005229	-0.174114	-0.174084	-0.171543	-0.005296	-0.005336
	6	0.144945	0.018992	0.019035	-0.018493	0.144768	0.144726
	7	-0.028692	-0.285607	-0.285730	-0.238693	-0.028562	-0.028607
	8	-0.006022	-0.214281	-0.214295	-0.212761	-0.006018	-0.006063
	9	2.000000	1.793422	1.793411	1.797805	-0.042760	-0.042312
	10	-0.031764	-0.227952	-0.227948	-0.215925	-0.031775	-0.031816
	11	-0.003434	-0.195726	-0.195717	-0.180833	-0.003450	-0.003489
ɵ	2	-0.013383	-0.000869	-0.000867	-0.002595	-0.013406	-0.013402
(rad)	5	-0.014913	0.016229	0.016229	0.016201	-0.014972	-0.014961
	8	-0.015403	0.025673	0.025673	0.025623	-0.015447	-0.015434
P_{ij}	3-4	-0.016890	-0.034213	-0.034199	-0.077728	-0.016803	-0.016830
(p.u.)	6-7	0.087880	-0.048806	-0.048718	-0.088328	0.088202	0.088155

Table 3-6- Residual errors for in the different estimations, when P_{inj9} has a gross error.

Component		R_{MSE}	R_{MCC} (σ=5)	R_{RQE} (σ=5)	R_{MCC} (σ=0.1)	R_{RQE} (σ=5)
P_{inj} (p.u.)	100	-0.067395	-0.067418	-0.048828	-0.000095	-0.000117
	1	-0.116712	-0.116693	-0.129089	-0.000085	-0.000127
	2	-0.140686	-0.140658	-0.168100	-0.000080	-0.000124
	3	-0.177594	-0.177570	-0.220860	-0.000074	-0.000126
	4	-0.185797	-0.185783	-0.166129	-0.000023	-0.000061
	5	-0.168884	-0.168855	-0.166314	-0.000067	-0.000106
	6	-0.125953	-0.125910	-0.163438	-0.000177	-0.000219
	7	-0.256915	-0.257038	-0.210001	0.000130	0.000085
	8	-0.208259	-0.208273	-0.206739	0.000004	-0.000041
	9	-0.206578	-0.206589	-0.202195	-2.042760	-2.042312
	10	-0.196188	-0.196185	-0.184161	-0.000011	-0.000052
	11	-0.192292	-0.192283	-0.177399	-0.000016	-0.000055
ɵ (rad)	2	0.012513	0.012516	0.010787	-0.000023	-0.000019
	5	0.031142	0.031143	0.031114	-0.000059	-0.000048
	8	0.041076	0.041076	0.041026	-0.000043	-0.000030
P_{ij} (p.u.)	3-4	-0.017323	-0.017309	-0.060838	0.000087	0.000060
	6-7	-0.136686	-0.136598	-0.176208	0.000322	0.000275

In Renyi's quadratic Entropy, the situation is really similar to what was described to Correntropy, by instead of evaluate the error between the measured value and the estimated value, it evaluates the difference between an error and the error distribution.

Table 3-7 - State estimations provided by each criteria, with different PW sizes. Gross error in P_{inj3}.

Component		$\hat{Z}_{MSE}$	σ=5		σ =0.5		σ =0.1	
			$\hat{Z}_{MCC}$	$\hat{Z}_{RQE}$	$\hat{Z}_{MCC}$	$\hat{Z}_{RQE}$	$\hat{Z}_{MCC}$	$\hat{Z}_{RQE}$
P_{inj}	100	0.089321	0.088939	0.099516	0.110986	0.110921	0.111439	0.111442
(p.u.)	1	-0.217044	-0.216641	-0.225003	0.003826	0.003722	0.004309	0.004314
	2	-0.398837	-0.398372	-0.415842	-0.081131	-0.081254	-0.080634	-0.080628
	3	1.510290	1.510090	1.483435	-0.054653	-0.054224	-0.057194	-0.057214
	4	-0.026576	-0.027089	-0.014604	-0.004997	-0.004979	-0.004973	-0.004974
	5	-0.064601	-0.064923	-0.063415	-0.005351	-0.005360	-0.005285	-0.005285
	6	-0.007997	-0.007895	-0.032027	0.144624	0.144551	0.144792	0.144796
	7	-0.162571	-0.162524	-0.132192	-0.028687	-0.028663	-0.028539	-0.028540
	8	-0.245947	-0.245571	-0.244202	-0.006243	-0.006291	-0.005979	-0.005976
	9	-0.248234	-0.247940	-0.244847	-0.042971	-0.043009	-0.042745	-0.042743
	10	-0.145301	-0.145354	-0.137610	-0.031880	-0.031891	-0.031756	-0.031755
	11	-0.082503	-0.082720	-0.073209	-0.003523	-0.003522	-0.003436	-0.003436
θ	2	-0.002179	-0.002149	-0.003081	-0.013344	-0.013332	-0.013416	-0.013417
(rad)	5	0.023917	0.023913	0.024172	-0.014841	-0.014812	-0.014994	-0.014995
	8	0.021950	0.021953	0.022203	-0.015317	-0.015289	-0.015469	-0.015470
P_{ij}	3-4	0.537890	0.538355	0.510286	-0.016278	-0.016211	-0.016889	-0.016892
(p.u.)	6-7	0.119419	0.119276	0.093654	0.088233	0.088196	0.088198	0.088200

Table 3-8 - Residual errors in the different estimations when P_{inj3} has a gross error.

Component		R_{MSE}	σ=5		σ=0.5		σ=0.1	
			R_{MCC}	R_{RQE}	R_{MCC}	R_{RQE}	R_{MCC}	R_{RQE}
P_{inj}	100	-0.022148	-0.022530	-0.011953	-0.000483	-0.000548	-0.000030	-0.000027
(p.u.)	1	-0.221369	-0.220966	-0.229328	-0.000499	-0.000603	-0.000016	-0.000011
	2	-0.318212	-0.317748	-0.335217	-0.000506	-0.000629	-0.000009	-0.000003
	3	-0.469710	-0.469910	-0.496565	-2.034653	-2.034224	-2.037194	-2.037214
	4	-0.021620	-0.022134	-0.009649	-0.000041	-0.000024	-0.000018	-0.000019
	5	-0.059372	-0.059693	-0.058186	-0.000122	-0.000131	-0.000056	-0.000055
	6	-0.152942	-0.152840	-0.176972	-0.000321	-0.000394	-0.000153	-0.000149
	7	-0.133879	-0.133832	-0.103500	0.000005	0.000029	0.000153	0.000152
	8	-0.239925	-0.239549	-0.238180	-0.000221	-0.000269	0.000043	0.000046
	9	-0.205456	-0.205162	-0.202068	-0.000193	-0.000230	0.000034	0.000035
	10	-0.113538	-0.113591	-0.105846	-0.000117	-0.000127	0.000008	0.000008
	11	-0.079069	-0.079286	-0.069775	-0.000089	-0.000088	-0.000002	-0.000002
θ	2	0.011204	0.011234	0.010302	0.000039	0.000051	-0.000034	-0.000034
(rad)	5	0.038831	0.038826	0.039085	0.000072	0.000101	-0.000081	-0.000082
	8	0.037353	0.037357	0.037607	0.000086	0.000114	-0.000065	-0.000066
P_{ij}	3-4	0.554780	0.555245	0.527176	0.000612	0.000679	0.000001	-0.000002
(p.u.)	6-7	0.031539	0.031396	0.005774	0.000353	0.000316	0.000318	0.000320

The same study was made, this time, including the outlier in the injected power of bus 3, to see its effluence in the line 3-4 power flow. The measurement vector was the same as in table 3-2- Measurement affected with noise, but assuming P_{inj_3}=1.98, introducing a large error. The results for this study can be found in table 3-7 and 3-8.

Comparing the tables 3-6 with 3-8, for a Parzen Window with size 5 or MSE, is possible to observe that the error in line 3-4 power flow increases when the gross error was introduced in one of the adjacent components, in this case bus 3. However, as que Parzen Window gets fitter the error in line 3-4 power flow decreases, in both cases (gross error is in P_{inj3} and in P_{inj9}). What was surprising here was the fact that, for σ=0.1, the residual in line 3-4 power flow was smaller when the gross error was introduced in P_{inj3}. This result shows that, although there is a gross error in an adjacent measurement, doesn't imply that the error in that line power flow has to increases.

In tables 3-7 and 3-8 were introduced another intermediate value to the Parzen Window size, σ=0.5. It's visible, for both MCC and RQE, that as σ decreases the estimated values approach gradually to the measured ones, ignoring the presence of outliers. When the Parzen Windows size is, at least, around ten times smaller than the error, that component losses its impact in the objective function evaluation and it's ignored. However, when it decreases too much it starts to ignore valid measurements and losses it significance, conducting to a bad result. This phenomenon is possible to observe in table 3-9, where the Parzen Windows size is 0.01. Using the Correntropy metric some of the estimations are far from the measurements and others are really close to them. The components that got a small residual were the ones possible to consider with an error that fits inside that little Parzen Window. Thus, we have some estimations totally adapted to some measurements and a lot of good measurements ignored, resulting in high residuals for those components.

Table 3-9- Estimations and residual errors for a too small Parzen Window.

Component		**Z**	**σ =0.01**			
			$\hat{Z}_{MCC}$	$\hat{Z}_{RQE}$	R_{MCC}	R_{RQE}
P_{inj} (p.u.)	100	0.111469	0.111465	0.111442	-0.000004	0.022121
	1	0.004325	0.004323	0.004314	-0.000001	0.221358
	2	-0.080625	-0.327784	-0.080628	-0.247159	0.318209
	3	1.980000	1.022480	-0.057215	-0.957520	-1.567505
	4	-0.004955	1.693474	-0.004974	1.698430	0.021602
	5	-0.005229	-2.466244	-0.005285	-2.461015	0.059317
	6	0.144945	0.803403	0.144796	0.658458	0.152793
	7	-0.028692	0.553882	-0.028540	0.582574	0.134031
	8	-0.006022	-5.580892	-0.005976	-5.574870	0.239970
	9	-0.042779	4.221090	-0.042743	4.263869	0.205491
	10	-0.031764	-0.031764	-0.031755	0.000000	0.113546
	11	-0.003434	-0.003434	-0.003436	0.000000	0.079067
ϴ (rad)	2	-0.013383	-0.013420	-0.013417	-0.000037	-0.011238
	5	-0.014913	-0.014913	-0.014995	0.000000	-0.038913
	8	-0.015403	-0.015403	-0.015470	0.000000	-0.037420
P_{ij} (p.u.)	3-4	-0.016890	-0.016890	-0.016892	0.000000	-0.554782
	6-7	0.087880	0.087880	0.088200	0.000000	-0.031218

Using the Renyi's quadratic entropy, all the residuals have grown in comparison with the case where σ=0.1. This can be easily explaining by the RQE perspective when applied to the error distribution. In fact, RQE aims to find an error distribution with minimum informational content. However, by decreasing too much the PW size, all the components are seen as outliers, since the error distribution doesn't fit inside the PW, as it is observable in table 3-9.

Once again, the residuals distributions are presented in the following charts. The residuals distributions were calculated though the PW method (with a standard deviation of 0.01) for the case with a gross error in $Pinj_3$. The error distributions were calculated to the MSE, MCC and RQE results, where the PW size applied to the optimization functions MCC and RQE assumed the size 0.1.

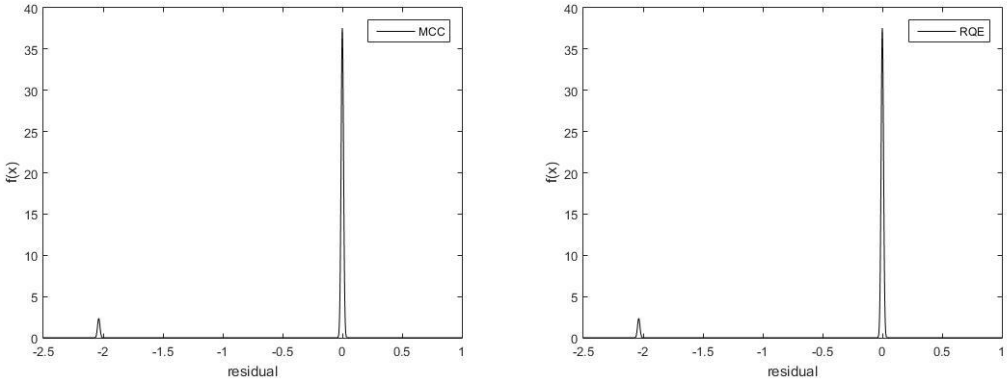

Figure 3-4 – Residual errors distribution for MCC, on the left , and RQE , on the right. Situation where the measurements had one gross errors.

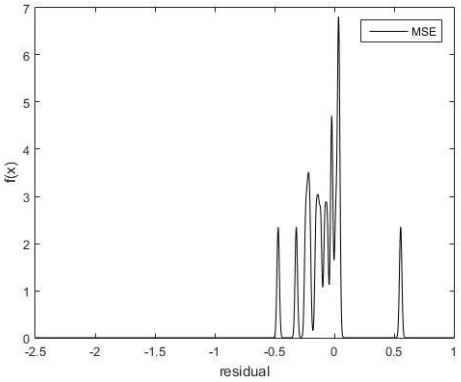

Figure 3-5 – Residual errors distribution for MSE. Situation where the measurements had one gross errors.

3.4.1 - Results analysis

In [31] had already been studied the behavior of MSE and MCC in situations with large errors, proving that MSE presents a bad behavior in that condition. In this dissertation was also possible to find equivalent results. Introducing large errors to the measurements, all the MSE estimations move away from the true measured value, being those estimations infected by the outlier. This effect is possible to observe in figure 3-5 where the error distribution assumes some peaks deviated from the zero value while, for MCC and RQE (figure 3-4) the distribution

assumes a Gaussian form centered in zero and then a peak located on the value of the gross error.

The fact that, in figure 3-4, the residuals distributions assume such form denotes that the estimation is computed ignoring the gross error. The isolated peak, far from the Gaussian, means that the outlier was identify and the estimation approached to the true value, generating a high residual for that component. Despite the big residual, we know that the estimated value was close to the real component value, as it was desirable.

The introduction of entropy based concepts had the objective of ignore the components affected with large error, taking advantage of the Parzen window size, and maintain the estimations free from the outliers effect.

Since the Parzen Window is adjusted conveniently, Correntropy shows a good behavior when facing an outlier, as shown before. What happens is that, when the residual term - $(z_i - \hat{z}_i)^2$ - is high, its magnitude is even more pronounced by the division of a Parzen Window (with size less than one). When applying the exponential, $e^{-\frac{(z_i-\hat{z}_i)^2}{2\sigma^2}}$, that component's result will be really small. Thus, can be said that this component is ignored and doesn't influence in the objective function, allowing all the other components to conduct to the optimal point. This particularity of MCC, change its behavior acting as different metrics depending on the distance between two points, was already explained in chapter 2.1.6 -.

The Renyi's Quadratic Entropy also presents a good behavior in the presence of gross errors. However, its approach is different from Correntropy's. The objective of the RQE is to find an error distribution with minimum information, that is, small variance and centered nearby zero. A Dirac impulse localized in zero would be the perfect estimation. However, the measurements are already contaminated with small noise. If the PW decreases too much, most of the components are ignored and it leads to a bad state estimation.

As seen before, Renyi's entropy and Correntropy have a powerful property: they are capable to ignore outliers and compute the estimation without consider the bad data. It only requires a convenient definition of the Parzen Window size, in a way that it is small enough to discard gross errors but big enough to allow the consideration of measurements with small noise. Other way, it will start to ignore a lot of acceptable components and loses observability of the network.

When the Parzen Window is too high Correntropy's estimator leads to the same solution as WLS and RQE leads to a close one.

3.5 - Chapter conclusions

From the analysis of this chapter, there are two relevant conclusions. The first one is about the inconvenience of the Cauchy-Schwarz distance applied to state estimation. As it was already explained, it tries to find an estimation vector collinear with the measurement vector, conducting to a distant point. In fact, closer states can be available without no requirement of collinearity.

The second, and most important conclusion, is about the application of maximum Correntropy and minimum Renyi's quadratic entropy criterions as state estimation objective functions. When the measurements are not corrupted with bad data, these metrics presented results similar to the ones obtained with MSE. More important than that is the behavior of those metrics in the presence of bad data. In fact, MCC and RQE have the ability to identify and ignore outliers, providing a convenient estimation to all the components. Thus, one must say

that the application of MCC and RQE in the state estimation problem has the advantage to provide a robust state estimator with no requirement of data pretreatment.

Chapter 4

Sensory fusion considering two sensor systems

Phasor Measurement Units have a unique ability to measure analogue voltage and current data in synchronism with a GPS-clock. Therefore, PMUs are considered to be important devices used as measure technology of power systems. Beyond other factors, due its ability to improve the precision of the state estimator by avoiding asynchronism errors, the implementation of this type of sensor has increased.

The integration of PMU in the electric power system leads to a situation with two distinct classes of sensors, which need to be combined. This combination can be done in different ways. The simplest, and most direct path, would be to integrate all the measurements in a vector and directly apply the WLS. However, there are some other possibilities. In this thesis, it is tested another procedure: the measurements are divided in two groups, according with the sensory system that gathered it. The weight assign to components with different origins varies, demonstrating more trust in one or another sensory system. This analysis is done in the framework of a multiple criteria modeling and Pareto-front discovery. The trust variation results in different values to the evaluation of each group of estimations, drawing a curve with axis: (fitness of group one, fitness of group two). The effect of the trust variation is analyzed to the different state estimation criterions, MSE, MCC and RQE.

In chapter 2.3.2 - data fusion strategies were exposed. This chapter aimed at proposing a sensory fusion method that combines conventional sensors with PMU. Both classes of sensors measure different properties, but their measurements complement each other, providing a better perception of the power system. Thus, through Boudjemaa, it is considered a fusion across attributes, and through Durrant-Whyte, a complementary fusion. The perspective of Dasarathy will be discussed later.

4.1 -Position of the sensors

The set of measurements were divided in two, in order to simulate the influence of two distinct classes of sensors, conventional and Phasor Measurement Units. In this way, in the next table is possible to see the origin of each measurement.

Table 4-1 - Measurements origin

	Bus/line	Sensor Class	Measured value
P_{inj} (p.u.)	100	Conventional (1)	0.111469
	1	Conventional (1)	0.004325
	2	PMU (2)	-0.080625
	3	Conventional (1)	-0.056763
	4	Conventional (1)	-0.004955
	5	PMU (2)	-0.005229
	6	Conventional (1)	0.144945
	7	Conventional (1)	-0.028692
	8	PMU (2)	-0.006022
	9	Conventional (1)	-0.042779
	10	Conventional (1)	-0.031764
	11	Conventional (1)	-0.000044
ϴ (rad)	2	PMU (2)	-0.013383
	5	PMU (2)	-0.014913
	8	PMU (2)	-0.015403
P_{ij} (p.u.)	3-4	Conventional (1)	-0.016290
	6-7	Conventional (1)	0.087880

The DC model will be used considering that there are PMU sensors on buses 2, 5 and 8. PMU measures voltage and current phasors. For the simplifications of DC model is considered that the voltage magnitude is 1 p.u. causing the injected power to assume the same value as the injected current. Thus, for the DC model will be consider that the PMU measurements are voltages phases and power injections (once for this model the power injection value is the same as current injection).

In other to obtain graphics with a better representation of the curve we want to expose, the measurements used were the ones already presented in chapter 3, but with a little deviation in one of the measurements, and it can be found in table 4-1.

4.2 - Trust variation among the different classes of sensors

The main idea of this section is to assign different weights to the fitness evaluations of each set of estimations, according to the sensory system that provided the measurement. For that, the measurement set was divided in two groups. Each group estimation errors were evaluated according to an evaluation metric, leading to two fitness evaluations. This evaluation metric is correspondent to the objective function of the state estimation. The total evaluation is obtained assigning an weight to each one of those evaluations as

$$fit = \alpha * fit_1 + \beta * fit_2 , \tag{4.1}$$

where:

- fit_1 and fit_2 corresponds to the fitness evaluation of residual errors, from group one and two, respectively. It can be calculated trough WLS, Correntropy or Renyi's quadratic Entropy;
- α and ß are the weight assign to the evaluation of each set of residual errors, where ß=α-1;

The optimal values assign to α and ß depend on multiple factors. Not only it depends on the confidence inherent to each sensory system, but also on the observability that the quantity of sensors and its position in the network give about the system.

In order to inspect the variation of α and ß influence, on the estimations, the state estimation was solved with assistance of EPSO algorithm. This time, the objective function is the maximization of (4.1) (expect using MSE, were was computed a minimization). The terms fit_1 and fit_2 were calculated according to each metric: least square, Correntropy and Renyi's quadratic entropy.

For the cases where the criterion of the state estimation is the MSE or MCC it's easy to decide how will be computed either fit_1 and fit_2, because each term are related with only one measurement and its estimation. The same doesn't happen with Renyi's Quadratic Entropy. For this measure each term is related with two residuals of two different measurements. As explain in chapter 3 the expression utilized to RQE is

$$H_2 = \sum_{i=1}^{n} \sum_{j=1}^{n} e^{-\frac{(r_i - r_j)^2}{2\sigma^2}} . \tag{4.2}$$

In this case, there are two possible ways to compute the fitness functions. The simplest way is to calculate each fitness function fit_1 and fit_2, relating each residual only with residuals from the same sensory system, respectively,

$$fit_1 = \sum_{i \epsilon n1} \sum_{j \epsilon n1} e^{-\frac{(\hat{\varepsilon}_i - \hat{\varepsilon}_j)^2}{2\sigma^2}} \text{ and} \tag{4.3}$$

$$fit_2 = \sum_{i \epsilon n2} \sum_{j \epsilon n2} e^{-\frac{(\hat{\varepsilon}_i - \hat{\varepsilon}_j)^2}{2\sigma^2}} . \tag{4.4}$$

The other way is to relate each residual with all the other residuals, independent of its origin, this option allows the sum of fit_1 with fit_2 to be equal with the fitness value of all the group together, through the following expressions

$$fit_1 = \sum_{i \epsilon n1} \sum_{j \epsilon n1 \wedge n2} e^{-\frac{(r_i - r_j)^2}{2\sigma^2}} \text{ and} \tag{4.5}$$

$$fit_2 = \sum_{i \epsilon n2} \sum_{j \epsilon n1 \wedge n2} e^{-\frac{(r_i - r_j)^2}{2\sigma^2}} . \tag{4.6}$$

In the last situation, what will introduce a difference between the optimal point found by the fusion process or the simple process (estimation considering only one set with all the measurements) is the values attributed to the parameters α and ß.

For the following tests, this cases will be denominated as Renyi's quadratic entropy 1 and 2, respectively.

4.2.1 - Influence of α and ß in both fitness functions

In order to ascertain the variation of parameters α and ß, tests were performed for α from 0 to 1 in intervals of 0.01, with ß=α-1. In the following graphics. 4.1 to 4.4, is possible to see the Pareto-front of non-dominated points, found by EPSO, for fit_1 and $fit_{2,}$ obtained by each of the different criteria.

These graphics illustrate the variation in both fitness functions, fit_1 and fit_2, according to the trust attributed to each group of measurements. In this way, when α assumes the maximum value, 1, only fit_1 has influence in the final fitness function. Thus, it aims to maximize fit_1 without paying any attention to the fit_2 value, in fact, ß assumes the value 0. Hence, it's easy to understand that fit_1 reaches its maximum when α is equal to 1. The same stands to fit_2, it reaches its maximum when ß=1 and α=0, once only fit_2 has influence in the final fitness function.

There is a particular situation that needs special attention. When using the Mean Squared Error and Correntropy, figures 4-1 and 4-2, the fitness function is evaluated depending on the residual error of each component estimation separately. Therefore, when ß=1 and α=0, the state estimator will perform estimations for all the 17 components bases only in 6 measurements, 3 of injected power and 3 of voltage phases. Obviously the system has no observability and, despite fit_2 reaches its maximum, fit_1 will decrease drastically. This happens because all the measurements will be estimated having in consideration only few measurements. Hence, the estimation for the group two will adjust perfectly with a minimum error, but the error in the estimations of group one will be high. This is the reason why fit_1 decreases so much. Taking this fact in consideration, both figures 4-1 and 4-2 show a graphic where the point (α, ß) = (0,1) appear, in the left, and other graphic without that point, in the right.

As it its visible in figures 4-1 and 4-2, on the right, for α is different than 0, is possible to observe gradual increase of fit_1 and, at the same time, a decrease of fit_2 as α increases.

for RQE as SE criterion (case 1), figure 4-3 on the left, is possible to observe two isolated points. In fact, it considers the relation of all the residuals within each group of sensors separately. For the comparison of image 4-3 with 4-1 and 4-2, can be concluded that RQE (case 1) is more susceptible to lose the system observability. In the left is possible to see those isolated points, while in the right these points ((α, ß) = (0,1) and (α, ß) = (1,0)) were excluded allowing a better perspective of the evolution of the fitness evaluations of each sensory system varying α and ß.

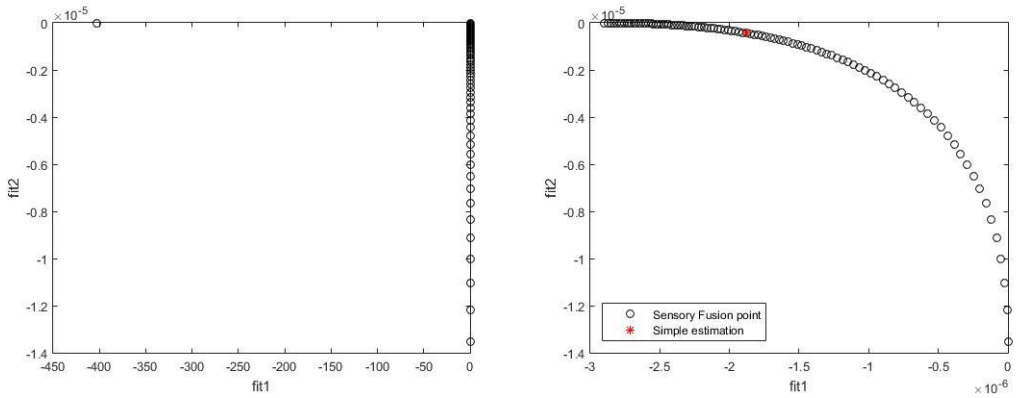

Figure 4-1 - Fitness 1 and 2 variation with α, assuming mean squared error. Representation of the simple fusion by the red dot

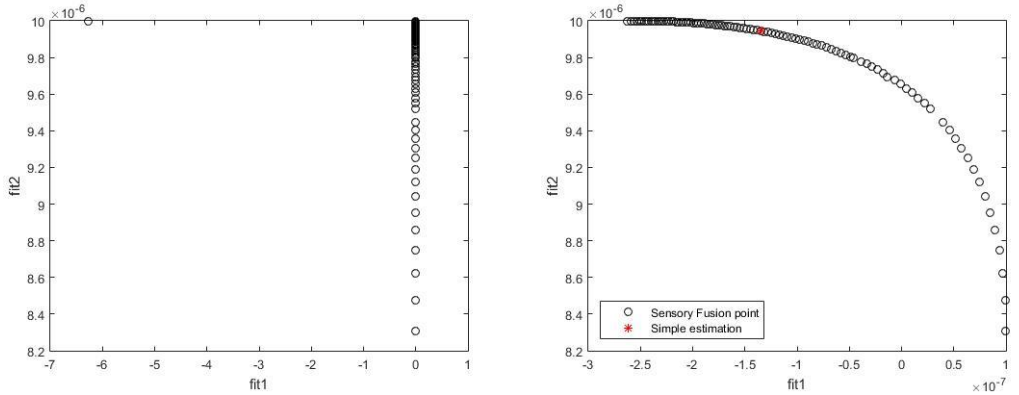

Figure 4-2 - Fitness 1 and 2 variation with α, assuming Correntropy. Representation of the simple fusion by the red dot

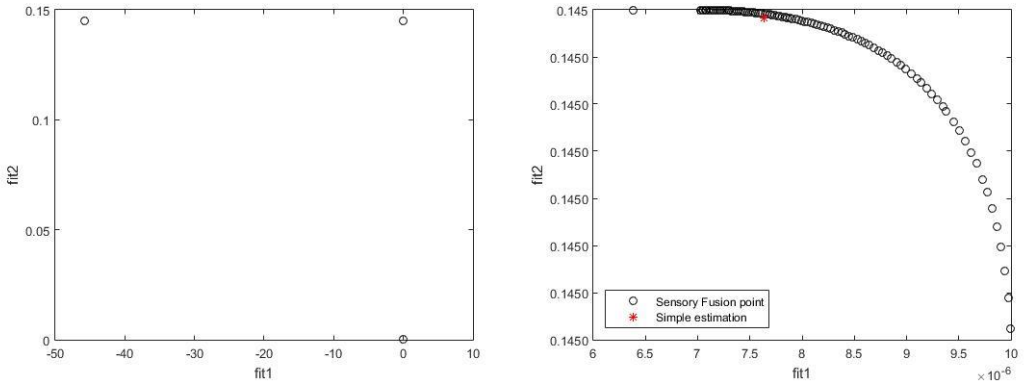

Figure 4-3 - Fitness 1 and 2 variation with α, assuming RQE case 1. Representation of the simple fusion by the red dot

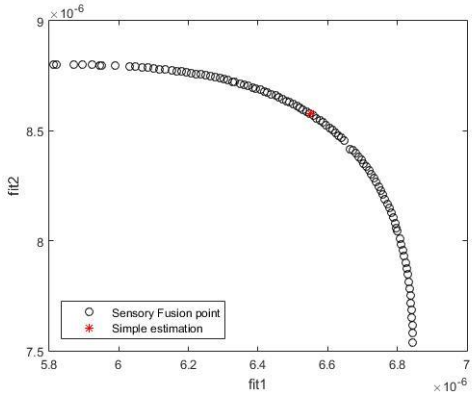

Figure 4-4- Fitness 1 and 2 variation with α, assuming RQE case 2.Representation of the simple fusion by the red dot.

The same distant point (α, ß) = (0, 1) for Renyi's quadratic entropy (case 2) is not visible and it's easy to justify why. In Renyi's quadratic entropy (case 2) the fit evaluation doesn't depend on the error of each measurement separately, but in its dimension comparing to all the other errors, as is can be seen by the equation (4.5) and (4.6).

In figures 4-1 to 4-4 is also possible to see the point equivalent to the simple fusion- the red dot. That is, the estimation considering only one set with all the measurements, where all

the measurements have the same weight to the final fitness function. As expected, the optimal point obtained with the simple fusion is situated among the curve of non-dominated points.

4.2.2 - Residual analysis

In the figures 4-5 to 4-11 is possible to observe the evolution of the residual errors, from the diverse estimations, in order to α. As explained before, for MSE and MCC criterions, when α=0 only the measurements of class two are considered and that number of measurements are insufficient to guarantee the system observability. In figures 4-1 and 4-2 is possible to observe an isolated point for α=0, now the same point is also observed in 4-5 and 4-6 represented by high errors in comparison with the errors for other values of α. In fact, these errors on estimations of group one is at least 100 times bigger in comparison with the errors for the other values of α. For a better observation of the evolution of the error, through the variation of α, the point α=0 was ignored. It is possible to observe a new graphic with representation of the evolution of the errors in 4-5(on the right side) to MCC and 4-6(on the right side) for MSE.

Also in figure 4-3 is possible to observe isolated points for the REQ (case 1). In figures 4-7 and 4-9(left side) is also possible to see these same points represented now, not by the deviation in the fitness functions, but by the deviation in the errors. The point with lower value for the fitness evaluation of the group one is represented by the high error presented in figure 4-7, for α=0. The point represented in figure 4-3 with lower value for the fitness evaluation of the group 2 is now represented in figure 4-12 for α=1, where is possible to observe errors extremely higher than the others obtain for different values of α.

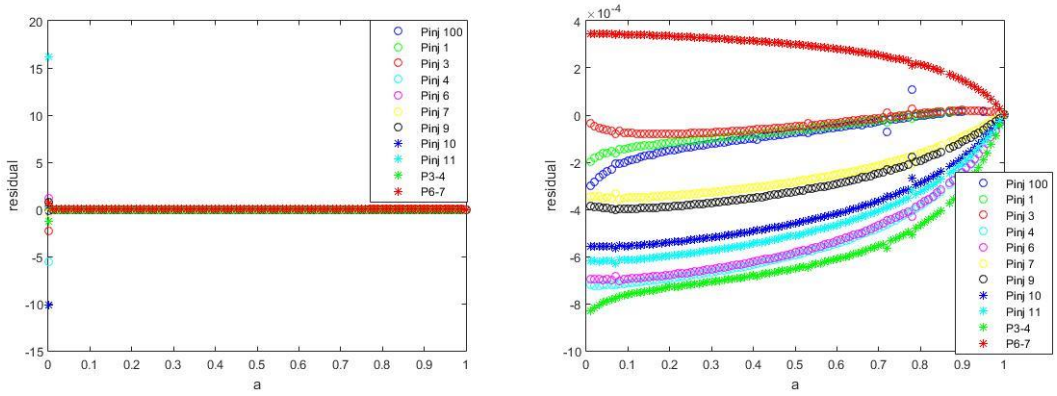

Figure 4-5- Residual errors (p.u.) for the group one using Correntropy, varying α. On the right side the point α=0 was ignored.

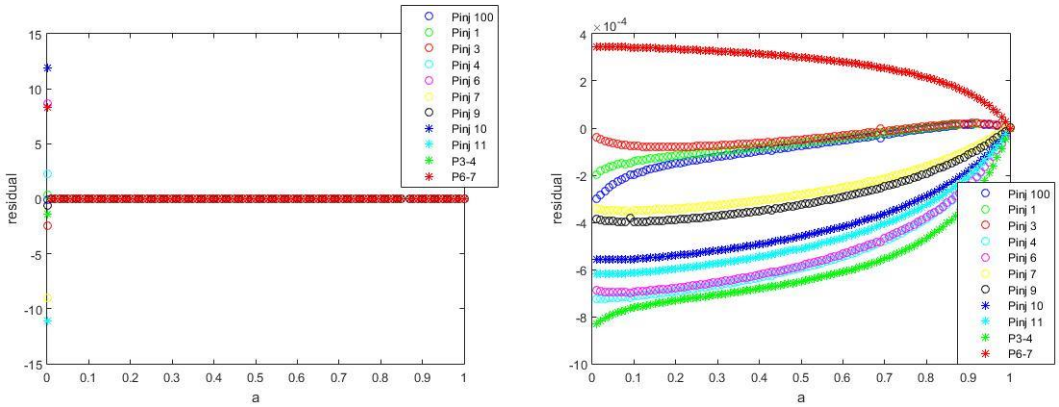

Figure 4-6- Residual errors (p.u.) for the group one using Mean Squared Error, varying α. On the right side the point α=0 was ignored.

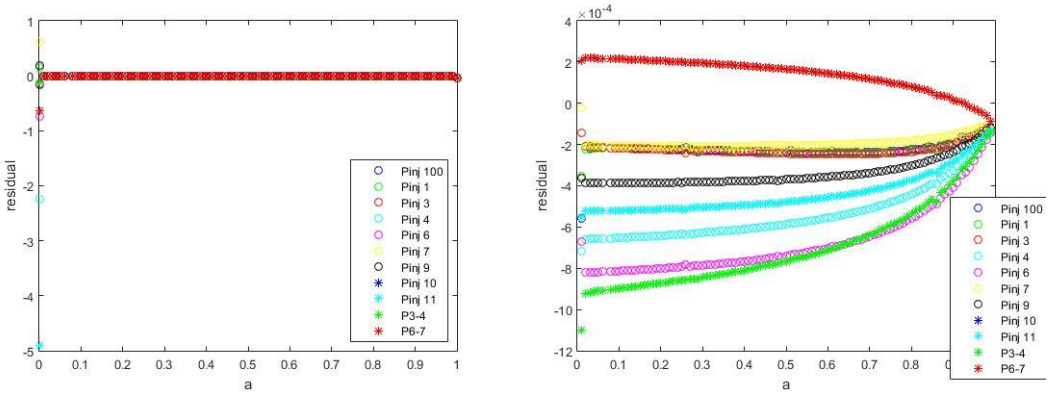

Figure 4-7 - Residual errors (p.u.) for the group one using Renyi's Quadratic entropy(case one), varying α. On the right side the point α=0 was ignored.

.

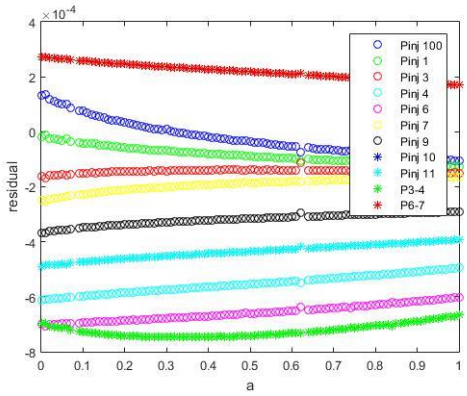

Figure 4-8 - Residual errors (p.u.) for the group one using Renyi's Quadratic Entropy (case 2), varying α.

For the second group of sensors only in the RQE case 1 were observed a case of an isolated point, with a residual error much higher than the others for different values of α. That situation is represented in figure 4-9. In the same figure but in the right side is represented the variation of the errors of the estimations for the group 2 in function of α, but ignoring the point α=1, or ß=0.

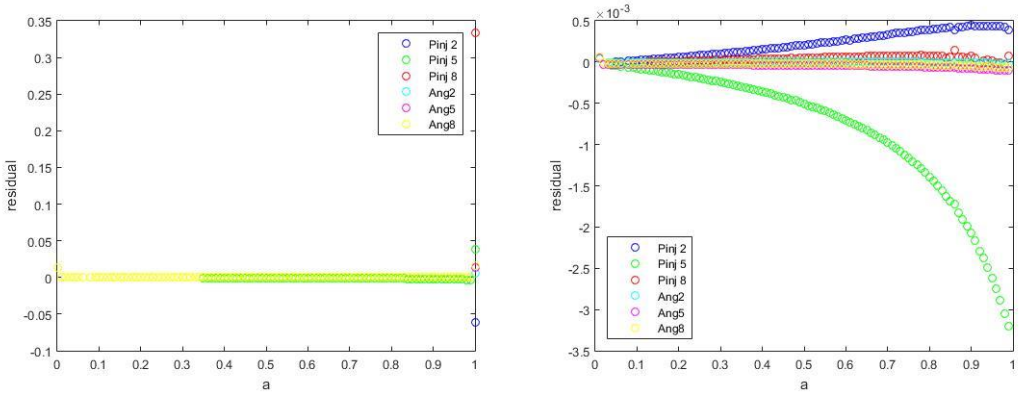

Figure 4-9- Residual errors (p.u.) for the group two, using Renyi's Quadratic entropy(case one), varying α. On the right side the point α=1 was ignored.

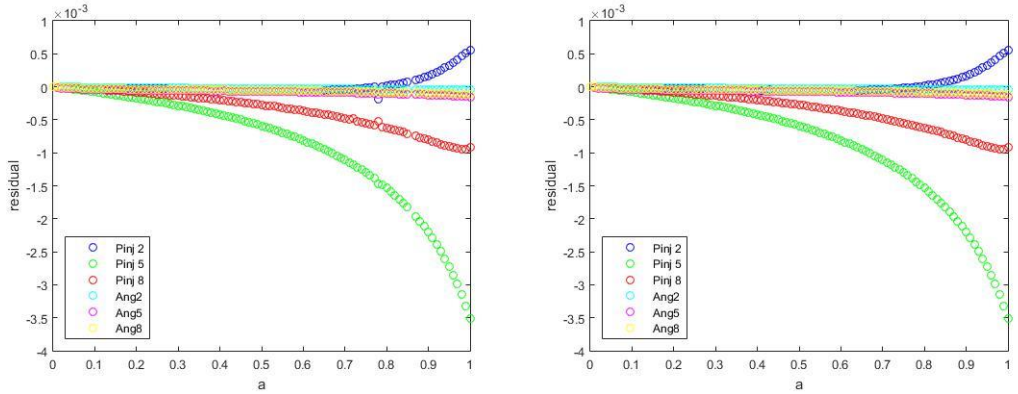

Figure 4-10-Residual errors (p.u.) for the group two, using MCC on the left and MSE on the right, varying α.

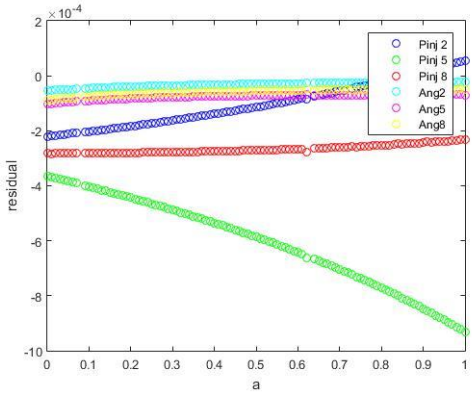

Figure 4-11-Residual errors (p.u.) for the group two, using Renyi's Quadratic Entropy (case 2), varying α

Is important to mention that, for MCC and MSE, despite what happens when α=0 to the group one, the same doesn't happen to the group two when ß=0. The errors in the estimations of group two increase as ß decreases (α increases), but there is not an isolated point with an superior dimension error, as it happens to the estimations of the group one when α=0. That's easily explain, because the group one has much more measurements and, even assuming that the measurements of the group 2 are in fault (that's what happen when it's assumed that ß=0, to MCC and MSE) the system is still observable. Hence, despite the growing of estimation errors, it doesn't grow so much as in the previous case, as it is visible by the point α=1 in figures 4-10.

As it was expected, in a general mode, the estimation errors of the group one, figures 4-5 to 4.8, decreases as α increases. Which is obvious, α determines the weight of that components to the objective function, once it grows the trust in that measurements increases and the estimations tends to approach their values. The same can be said to the measurements of the group two, but when α decreases, that is, ß increases, figures 4-9 to 4-11.

Using the minimization of MSE or Correntropy maximization as SE objective function, it's possible to see a big slope of the residuals through the parameter α. However, when using the Renyi's Quadratic Entropy (case 2) that slope isn't that much accentuated, figure 4-11. In fact, the computation of the Renyi's Quadratic Entropy (case 2) was applying the expression (4.5) and (4.6). Thus, the fit_1 is given by the sum of all the components from group one interacting with all the other errors, even with components from the group two. In this way, even when α=0 the system doesn't loss its observability, because the components from the group one are still taken in consideration, even that the relations with the errors inherent to the group two are prevalent.

In RQE (case 1) the fitness function is calculated considering only residuals referring to the same group. As it correlates all the residuals of the same group, to calculate each fitness evaluation, it promotes much more interaction by the errors obtained by each sensory system. In this way, this criterion is more susceptible to lose the system observability, while in case 2, measurements provided by both systems interact with each other maintaining the system observability.

4.3 - Chapter conclusions

The definition of α and ß parameters conducts the state estimation either closer to a measurement set or to the other, depending on the trust attributed to each sensory system. It's important to mention that, excluding the cases where the system loses its observability, each one of the non-dominated points, present in figure 4-1 to 4-4, are possible. The question is: which point should be assumed as the real state?

One proposal is to find stochastic weights for the parameters α and ß. This can be done in a future work, requiring lot of research and tests. Such investigation must evaluate the probability of each sensory system to conduct the state estimation to the real one, depending on: sensor type, quantity, location, synchronism and other possible factors.

As that information is not yet available, another way to find an optimal point has to be explored. The following chapter is dedicated to that search, in order to find an optimal fusion point.

The sensory fusion method presented in this chapter combines conventional sensors with PMU. Both types of sensors measure different properties, but their measurements complement each other, providing a better perception of the power system state. Thus, this sensory fusion method can be classified as complementary, by the Durrant-Whyte definition, and fusion across attributes, by Boudjemaa. As it is possible to observe in the previous section, the variation of each sensory system weight originates a Pareto-front, resulting in a decision problem. Through a proper definition about these weights, the decision of the optimal fusion state is done automatically. That is, the fusion of the input features generates an output decision. Thus, the proposed stochastic fusion is, through Dasarathy, a fusion with input/output characteristics of the type feature input/ decision output, FeI-DeO.

Chapter 5

Selecting the optimal fusion point

In chapter 4, was concluded that the variation of each sensory system weight, on the SE objective function, originates a Pareto-front with the same form as 5.1. The extremal points conduct to situations where only one class of the sensors was considered, and it can be seen as less reliable situation since the system can losses its observability. However, approaching to a situation of α close to ß and close to 0.5 both fit_1 and fit_2 decrease but, all the measurements are used to perform a better result. That situation is somewhere represented in the graphic where both fit_1 and fit_2 assume high values.

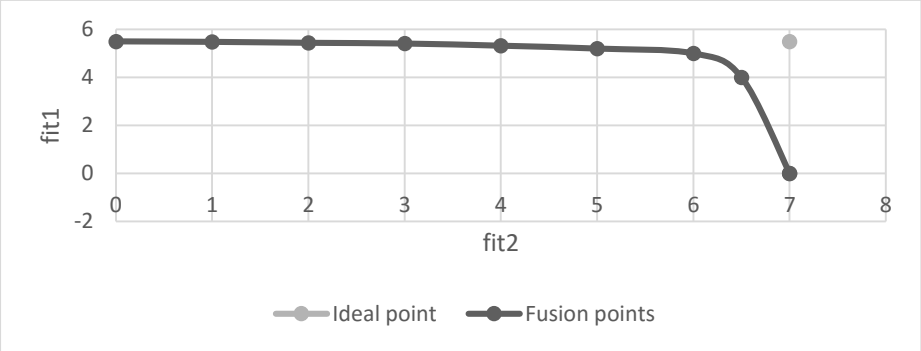

Figure 5-1- Typical Pareto-front obtained by the multiple fusion points, depending on α and ß, in black. Ideal point (maxfit1, maxfit2) represented by the grey point.

There isn't just a single way to decide which point represents better the true state of the power system. Still, one simple way to choose an optimal point can be select the one closest to the ideal point given by (Max_{fit1}, Max_{fit2}).

The values for Max_{fit1} and Max_{fit2} are easy to obtain with no need to draw all the curve. In fact, when only one class of sensor is considered, α=1 or ß=1, only its fitness function is maximized, ignoring the measurements provided by the other sensor system. Hence, when α is equal to 1 the maximum value of fit_1 is reached, and when ß is equal to 1 the maximum point of fit2 is achieved.

Another issue is how to measure that distance. Which is the metric that leads to a minimal error? In order to choose a metric that better approaches the optimal to the ideal point, some tests were provided, evaluating the effect of different metrics on that step of the sensory fusion. The explored metrics, applied to the search of the optimal point were:

- Absolute value norm- L1;
- Euclidian norm- L2;
- Correntropy Induced Metric-CIM;

At this point there are two different metrics simultaneously applied to the SE objective function:

- The evaluation metrics, responsible for the calculation of fit_1 and fit_2;
- The fusion metric.

The fusion metric is the one that measures the distance between the optimal point and the ideal one. In resume the data fusion will be done in two stages:

- Determination of the ideal point (max_{fit1}, max_{fit2});
- Search for the optimal fusion point closest to the ideal one;

In order to evaluate the impact of the different metrics in the data fusion, some different combinations between the fusion metrics and the evaluation metrics were analyzed. The evaluation metrics are the ones already explored in chapter 3 and 4:

- Minimum Mean squared error-MSE;
- Maximum Correntropy critetion - MCC;
- Minimum Renyi's quadratic entropy. For RQE, two approaches of its calculation were considered. First, the estimation errors just interact with errors within the same group, RQE1.Second, each error interacts with all the errors from both groups, RQE2. In the previous chapter these expressions are presented in (4.3) and (4.4) for RQE 1 and (4.5) and (4.6) for RQE2.

5.1 -Optimization function

Thus, the SE optimization function will suffer an adaptation. The new objective functions are presented in tables 5-1 to 5-3. The parameter σ refers to the PW size and n_a is the set of measurements provided by the sensory system a. N consists in the aggregation of measurements provided by both sensory systems, 1 and 2. $N=n_1 U n_2$.

That combinations were tested either without and with gross errors.

Table 5-1- Objective function using the metric L1 to find the closest point to the ideal one.

Objective function		**Fusion metric: L1**
Evaluation metric	MSE	$\min \left\|max_{fit1} - \sum_{i\epsilon n1}(z_i - \hat{z}_i)^2\right\| + \left\|max_{fit2} - \sum_{i\epsilon n2}(z_i - \hat{z}_i)^2\right\|$
	MCC	$\min \left\|max_{fit1} - \sum_{i\epsilon n1} e^{-\frac{(z_i-\hat{z}_i)^2}{2\sigma^2}}\right\| + \left\|max_{fit2} - \sum_{i\epsilon n2} e^{-\frac{(z_i-\hat{z}_i)^2}{2\sigma^2}}\right\|$
	RQE 1	$\min \left\|max_{fit1} - \sum_{i\epsilon n1}\sum_{j\epsilon n1} e^{-\frac{(r_i-r_j)^2}{2\sigma^2}}\right\| + \left\|max_{fit2} - \sum_{i\epsilon n2}\sum_{j\epsilon n2} e^{-\frac{(r_i-r_j)^2}{2\sigma^2}}\right\|$
	RQE 2	$\min \left\|max_{fit1} - \sum_{i\epsilon n1}\sum_{j\epsilon N} e^{-\frac{(r_i-r_j)^2}{2\sigma^2}}\right\| + \left\|max_{fit2} - \sum_{i\epsilon n2}\sum_{j\epsilon N} e^{-\frac{(r_i-r_j)^2}{2\sigma^2}}\right\|$

Table 5-2 - Objective function using the Euclidian metric to find the closest point to the ideal one.

Objective function		**Fusion metric: L2**
Evaluation metric	MSE	$\min \left(max_{fit1} - \sum_{i\epsilon n1}(z_i - \hat{z}_i)^2\right)^2 + \left(max_{fit2} - \sum_{i\epsilon n2}(z_i - \hat{z}_i)^2\right)^2$
	MCC	$\min \left(max_{fit1} - \sum_{i\epsilon n1} e^{-\frac{(z_i-\hat{z}_i)^2}{2\sigma^2}}\right)^2 + \left(max_{fit2} - \sum_{i\epsilon n2} e^{-\frac{(z_i-\hat{z}_i)^2}{2\sigma^2}}\right)^2$
	RQE 1	$\min \left(max_{fit1} - \sum_{i\epsilon n1}\sum_{j\epsilon n1} e^{-\frac{(r_i-r_j)^2}{2\sigma^2}}\right)^2 + \left(max_{fit2} - \sum_{i\epsilon n2}\sum_{j\epsilon n2} e^{-\frac{(r_i-r_j)^2}{2\sigma^2}}\right)^2$
	RQE 2	$\min \left(max_{fit1} - \sum_{i\epsilon n1}\sum_{j\epsilon N} e^{-\frac{(r_i-r_j)^2}{2\sigma^2}}\right)^2 + \left(max_{fit2} - \sum_{i\epsilon n2}\sum_{j\epsilon N} e^{-\frac{(r_i-r_j)^2}{2\sigma^2}}\right)^2$

Table 5-3 - Objective function using the Correntropy Induced Metric to find the closest point to the ideal one.

Objective function		**Fusion metric: Correntropy induced metric-CMI**
Evaluation metric	MSE	$\max \left(e^{-\frac{\left(max_{fit1} - \sum_{i\epsilon n1}(z_i-\hat{z}_i)^2\right)^2}{2\sigma^2}} + e^{-\frac{\left(max_{fit2} - \sum_{i\epsilon n2}(z_i-\hat{z}_i)^2\right)^2}{2\sigma^2}}\right)$
	MCC	$\max \left(e^{-\frac{\left(max_{fit1} - \sum_{i\epsilon n1} e^{-\frac{(z_i-\hat{z}_i)^2}{2\sigma^2}}\right)^2}{2\sigma^2}} + e^{-\frac{\left(max_{fit2} - \sum_{i\epsilon n2} e^{-\frac{(z_i-\hat{z}_i)^2}{2\sigma^2}}\right)^2}{2\sigma^2}}\right)$
	RQE 1	$\max \left(e^{-\frac{\left(max_{fit1} - \sum_{i\epsilon n1}\sum_{j\epsilon n1} e^{-\frac{(r_i-r_j)^2}{2\sigma^2}}\right)^2}{2\sigma^2}} + e^{-\frac{\left(max_{fit2} - \sum_{i\epsilon n2}\sum_{j\epsilon n2} e^{-\frac{(r_i-r_j)^2}{2\sigma^2}}\right)^2}{2\sigma^2}}\right)$
	RQE 2	$\max \left(e^{-\frac{\left(max_{fit1} - \sum_{i\epsilon n1}\sum_{j\epsilon N} e^{-\frac{(r_i-r_j)^2}{2\sigma^2}}\right)^2}{2\sigma^2}} + e^{-\frac{\left(max_{fit2} - \sum_{i\epsilon n2}\sum_{j\epsilon N} e^{-\frac{(r_i-r_j)^2}{2\sigma^2}}\right)^2}{2\sigma^2}}\right)$

The results of the measurement fusion without and with gross errors can be found in the next sections. The analyzed SE objective functions were here described and justified.

The measurement set, used to validate the model, was the same used in chapter 3 and can be found in the following table.

Table 5-4 - Measurements used to find the optimal point (with and without gross errors). Part one: active power.

Measured Values	**Bus/line**	**Without Gross errors**	**With a gross error**
P_{inj} (p.u.)	100	0.111469	0.111469
	1	0.004325	0.004325
	2	-0.080625	-0.080625
	3	-0.056763	1.980000
	4	-0.004955	-0.004955
	5	-0.005229	-0.005229
	6	0.144945	0.144945
	7	-0.028692	-0.028692
	8	-0.006022	-0.006022
	9	-0.042779	-0.042779
	10	-0.031764	-0.031764
	11	-0.003434	-0.003434

Table 5-5 - Measurements used to find the optimal point (with and without gross errors). Part two: Voltage phases and lines power flows.

Measured Values	Bus/line	Without Gross errors	With a gross error
ϴ (rad)	2	-0.013383	-0.013383
	5	-0.014913	-0.014913
	8	-0.015403	-0.015403
P_{ij} (p.u.)	3-4	-0.01689	-0.01689
	6-4	0.08788	0.08788

5.2 - Sensory fusion optimal point - Measurements without gross errors

For what was already explained in this chapter the values Maxfit1 and Maxfit2 are achieved when α=1 ∧ ß=0 and α=0 ∧ ß=1, respectively. Thus, it's easy to identify the ideal point without draw all the Pareto-front. In the next table is possible to consult the ideal points for each one of the evaluation metrics.

Table 5-6- Ideal point to the DC model without gross errors.

Metric	max_{fit1}	max_{fit2}
MSE	-7.53562572126791E-22	-7.26228534954962E-30
MCC	1.09999999999999E+01	6.00000000000000E+00
RQE1	2.52999999174548E+02	1.67999999684225E+02
RQE2	1.86999999354139E+02	1.01999999840997E+02

For a comparison of the state estimations obtained with the diverse objective functions present in table 5-1 to 5-3, in figures 5-2 to 5-5 is possible to see graphics with the residual errors for all the estimations, using the different fusion metrics L1 and L2. Once it doesn't matter if the error is positive or negative for these results analysis, the representation of the errors will be done through its absolute value. The following results were providing using the different evaluation metrics. When the evaluation metric is MCC or RQE the used PW was σ=2. The application of CIM as a fusion metric is a more delicate situation, thus, this metric will be discussed after the absolute and Euclidean norms.

In figures 5-2 to 5-5 the residual errors obtained for the minimum MSE and the MCC are really close. In fact, that is not a surprise, once the Parzen window's applied to the MCC was a large window with size 2, it must present values really close or equal to the ones obtained by the minimum MSE.

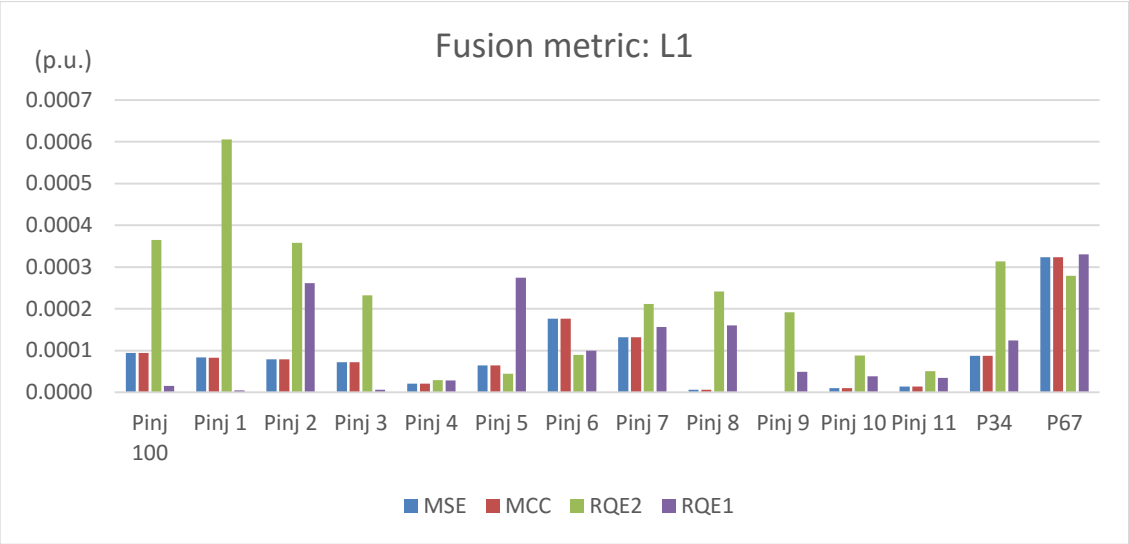

Figure 5-2 - Residual errors for each estimation of power injections and lines power flows, using L1 as the fusion metric and each one of the evaluation metrics.

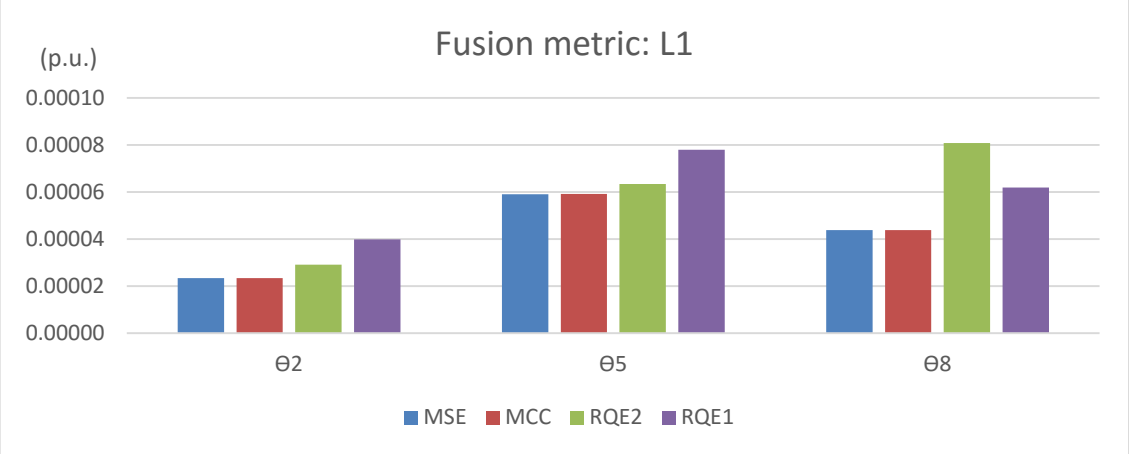

Figure 5-3 - Residual errors for each estimation of Voltage phase, using L1 as the fusion metric and each one of the evaluation metrics.

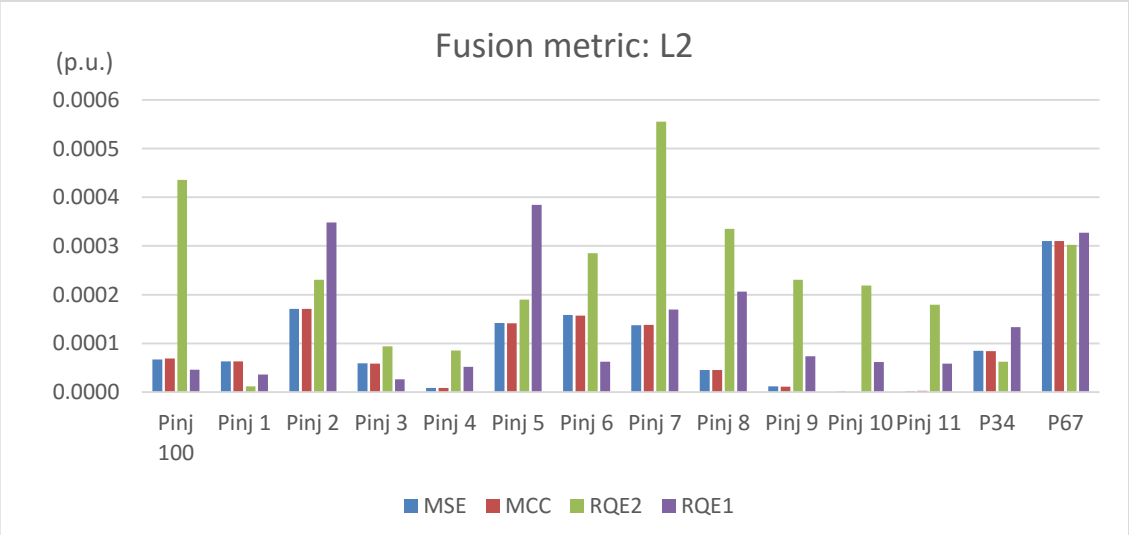

Figure 5-4 - Residual errors for each estimation of power injections and lines power flows, using L2 as the fusion metric and each one of the evaluation metrics.

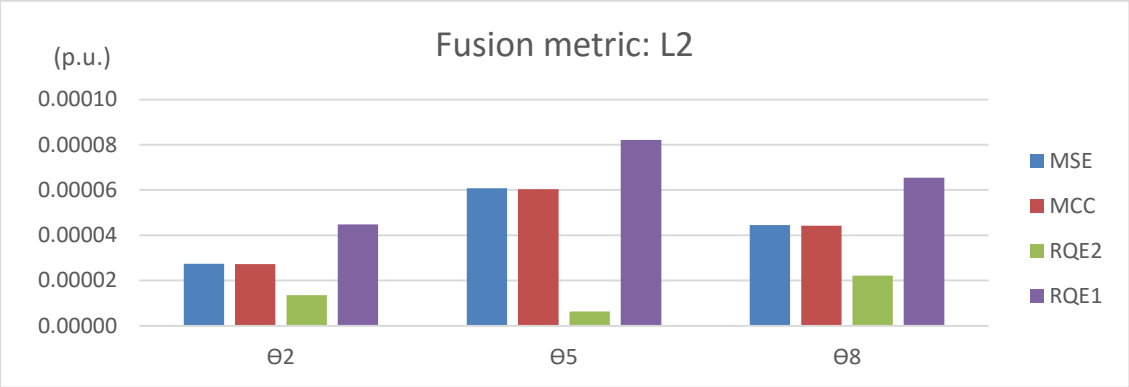

Figure 5-5 - Residual errors for each estimation of Voltage phase, using L2 as the fusion metric and each one of the evaluation metrics.

The residual errors for the power injections and lines power flows present better results for the evaluation metrics MSE and MCC in comparison with RQE, either with L1 or L2 as fusion metrics. For most of the components, the estimation errors provided by the minimum RQE2 (case 2) are higher than the ones obtained by the other evaluation metrics. RQE2 provided residual errors even bigger than the noise introduced in the original components values (power flow result without noise), that is, 0.0004 p.u. Despite RQE1 (case 1) shows in components higher errors than the ones obtained with the MSE and MCC, its errors are within the noise introduced in the original values. Thereby, the point obtained with the RQE case 1 is still a reliable one, despite its difference to the one obtained for MCC and MSE.

Because, when using the CIM as the fusion metric, its PW size has a big influence in the results, graphics with the errors provided by this metric are presented in a different form. Each of the following graphics is referred to a different evaluation metric and presents different residual errors for different values of the Parzen Window's size.

For a better comprehension of the results, in table 5-7, can be seen, not only the maximum value each of the fitness evaluation can reach, but also, the minimum. In this way it's possible to see which can be the maximum absolute distance between the ideal and the worst possible fusion point. It's important to mention that the selection of a Parzen Window's size doesn't depend on the maximum distance, but the minimum acceptable. If the Parzen Window is too small will start to ignore the optimal point closer to the ideal one.

Table 5-7 - Maximum and minimum values each fitness evaluation reaches for each group of sensors using the different metrics as objective function of the estimation.

	max_{fit1}	min_{fit1}	max_{fit2}	min_{fit2}
MSE	-7.536E-22	-1.052E+03	-7.262E-30	-7.684E-06
MCC	1.100E+01	3.892E+00	6.000E+00	6.000E+00
RQE 1	1.210E+02	5.550E+01	3.600E+01	3.582E+01
RQE2	1.870E+02	1.870E+02	1.020E+02	1.020E+02
	Max. difference for fit1		**Max. difference for fit2**	
MSE	1.052E+03		7.684E-06	
MCC	7.108E+00		9.562E-07	
RQE 1	6.550E+01		1.787E-01	
RQE 2	5.680E-08		8.078E-08	

As it is possible to see, using the CIM as the fusion metric, the optimal point depends on the parameter σ, PW size. When it is too high, figures 5-6 to 5-13, the errors start to grow. This is what happens from σ=5 to σ=20 and σ=100, the estimation errors start to grow,

independent of the evaluation metric, MSE, MCC and RQE. The definition of a proper PW size is complicated. For instance, for MSE the maximum difference between the ideal and the most distant point for the component fit_1 is around 1000 p.u.. However, the Parzen Window's size 100 or 20 is already damaging the results. Thus, select the Parzen Window's size considering the maximum acceptable difference is not a good practice, in this case.

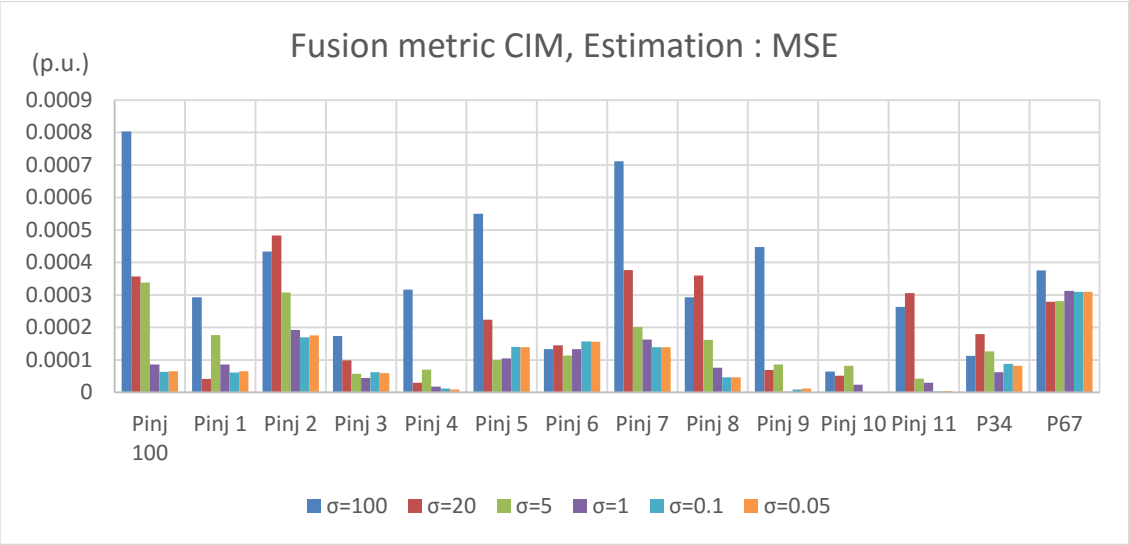

Figure 5-6 - Residual errors in buses power injections and line power flows, using CIM as the fusion metric (with different Parzen window's size)and MSE as evaluation metric.

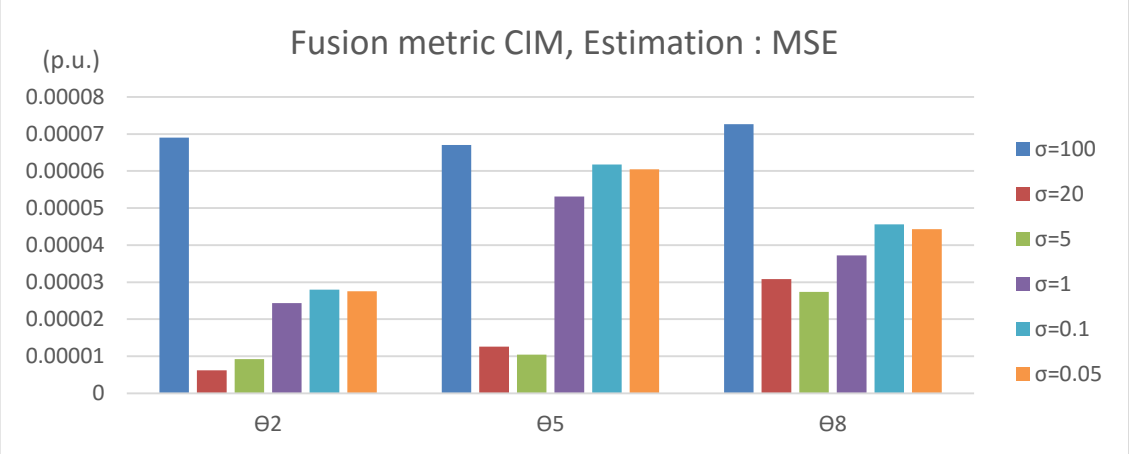

Figure 5-7- Residual errors in voltage phases, using CIM as the fusion metric (with different Parzen window's size)and MSE as evaluation metric.

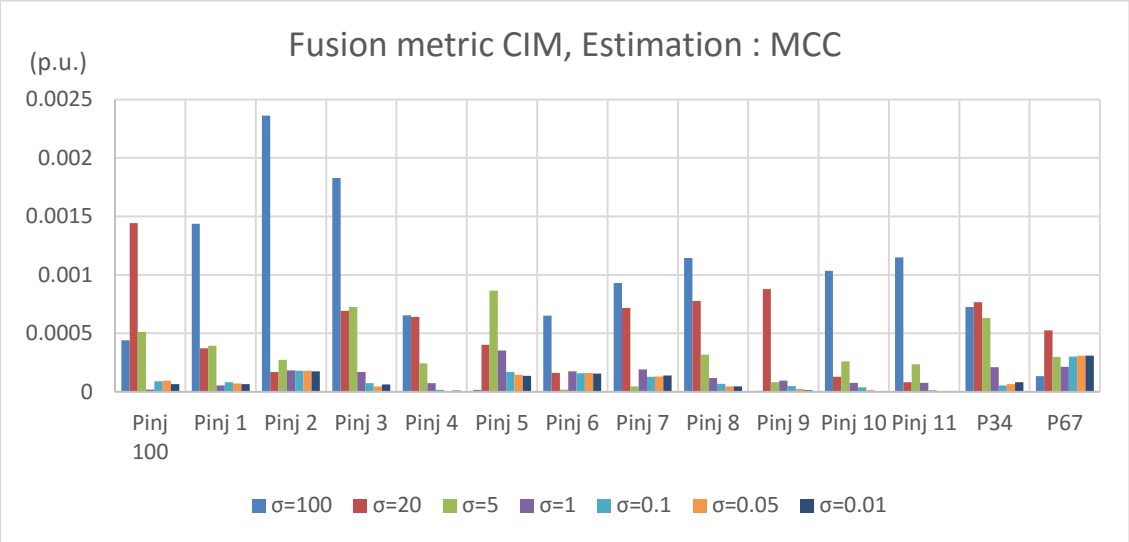

Figure 5-8 - Residual errors in buses power injections and line power flows, using CIM as the fusion metric (with different Parzen window's size)and MCC as evaluation metric.

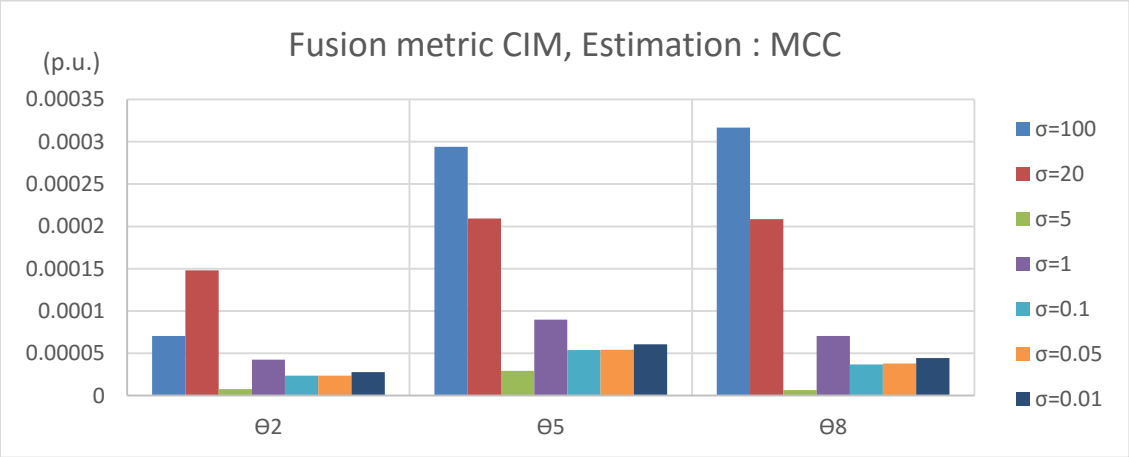

Figure 5-9 - Residual errors in voltage phases, using CIM as the fusion metric (with different Parzen window's size)and MCC as evaluation metric.

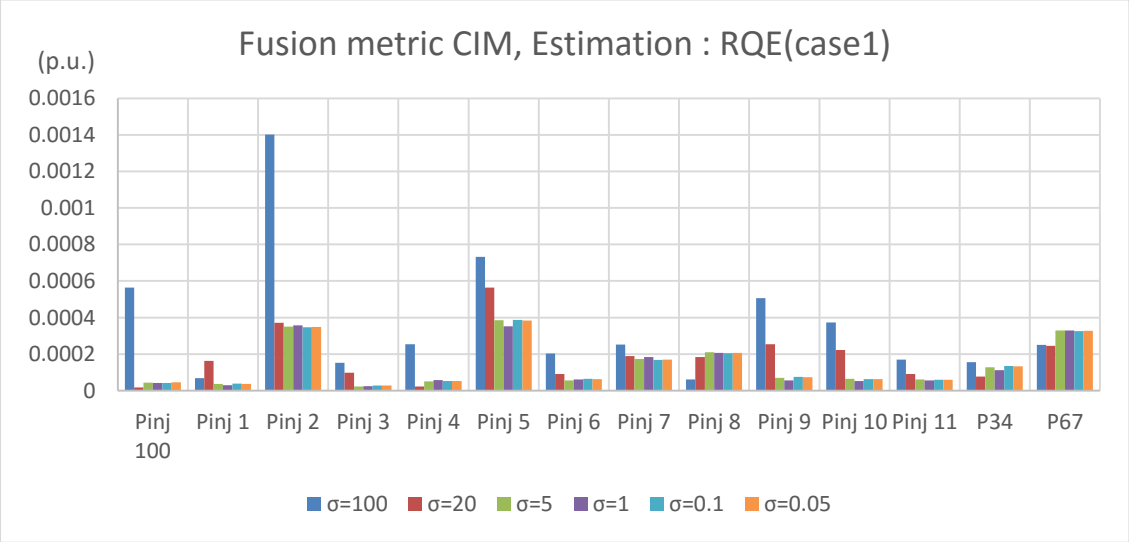

Figure 5-10 - Residual errors in buses power injections and line power flows, using CIM as the fusion metric (with different Parzen window's size)and RQE (case 1) as evaluation metric.

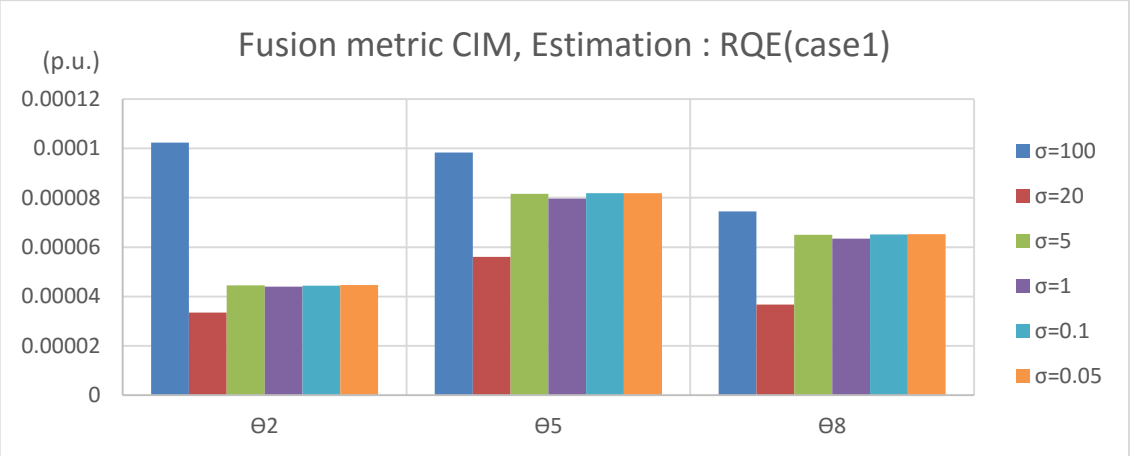

Figure 5-11 - Residual errors in voltage phases, using CIM as the fusion metric (with different Parzen window's size)and RQE (case 1) as evaluation metric.

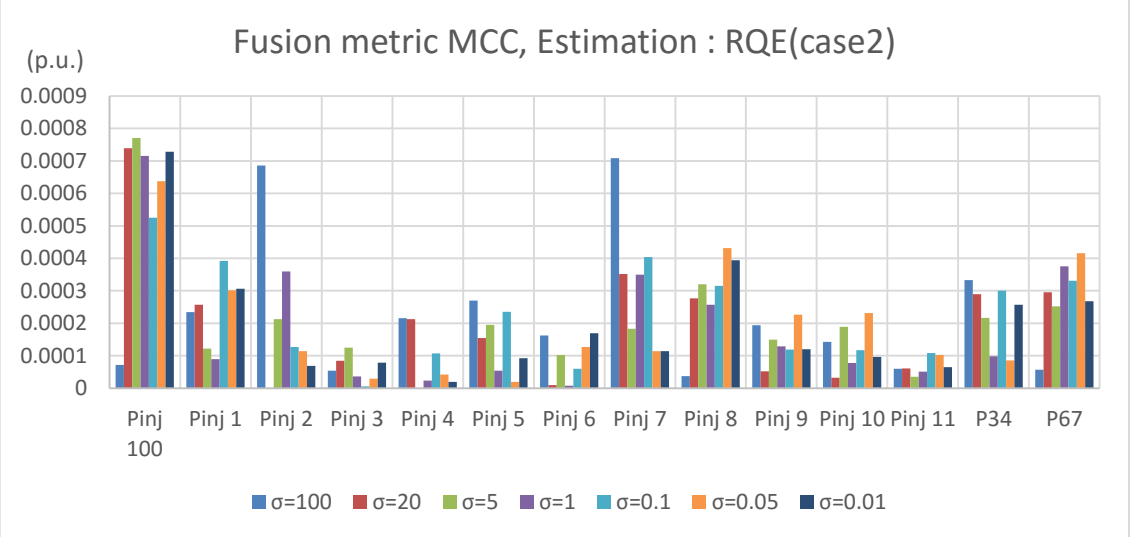

Figure 5-12 - Residual errors in buses power injections and line power flows, using CIM as the fusion metric (with different Parzen window's size)and RQE (case 2) as evaluation metric.

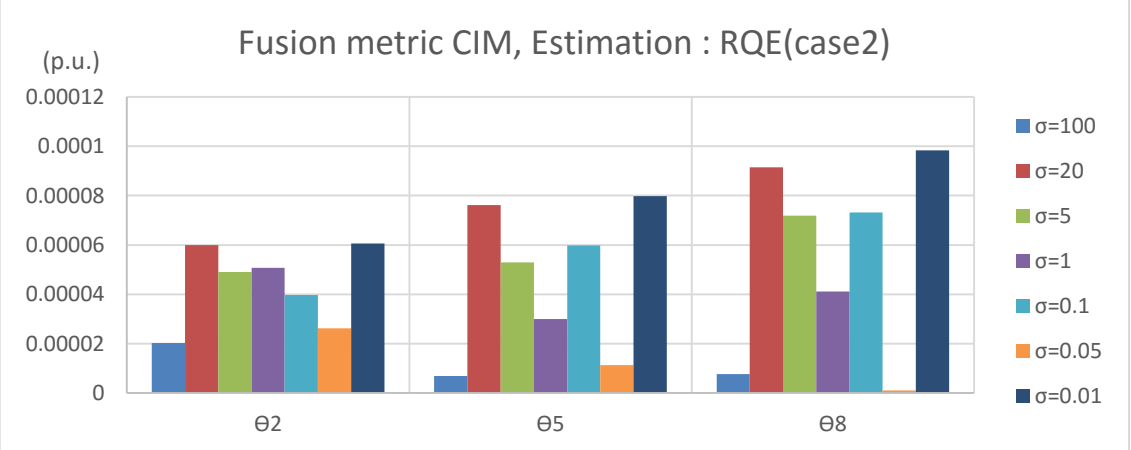

Figure 5-13 - Residual errors in voltage phases, using CIM as the fusion metric (with different Parzen window's size)and RQE (case 2) as evaluation metric.

Despite PW with high size can lead to bad results, it is necessary to pay attention when the Parzen Window's size is too small. In fact, when it is too low it starts to ignored even the optimal point. This was what happened for Parzen Window's size 0.0001, where, for all the evaluation metrics the estimation errors reached really high values, figure 5-14. What happened in this case was that the fitness evaluation ranges of group one and group two are not the same, thus, by decreasing the Parzen Windows's size it starts to ignore the measurements of one group of the sensors and consider only the other, where the variation of the fitness evaluation is smaller. That's why some of the elements present estimations with really big errors and others with really small, for σ=0.0001.

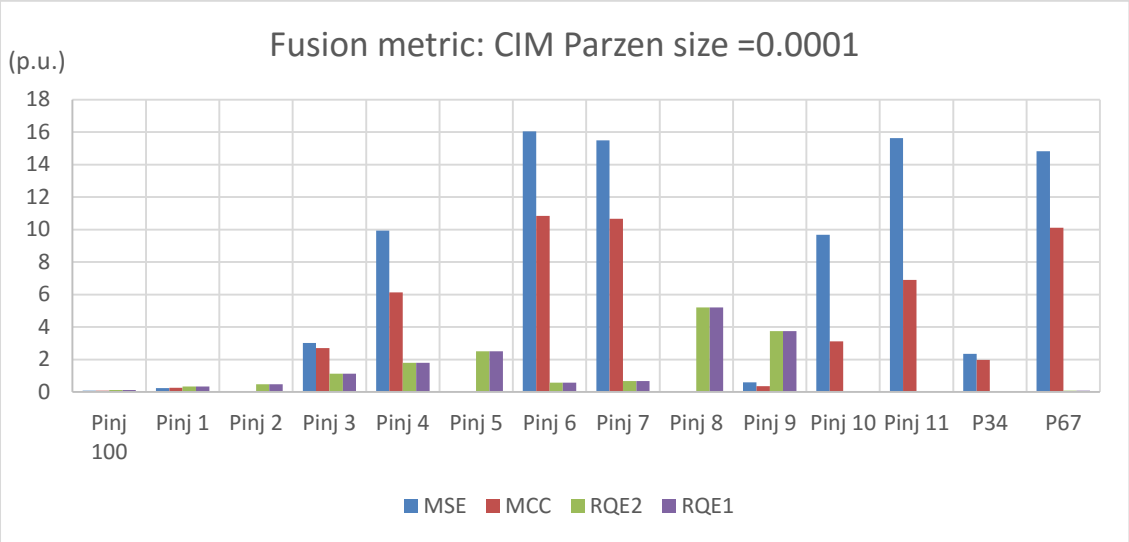

Figure 5-14 - Residual errors in buses power injections and line power flows using CIM as the fusion metric, with Parzen size 0.0001.

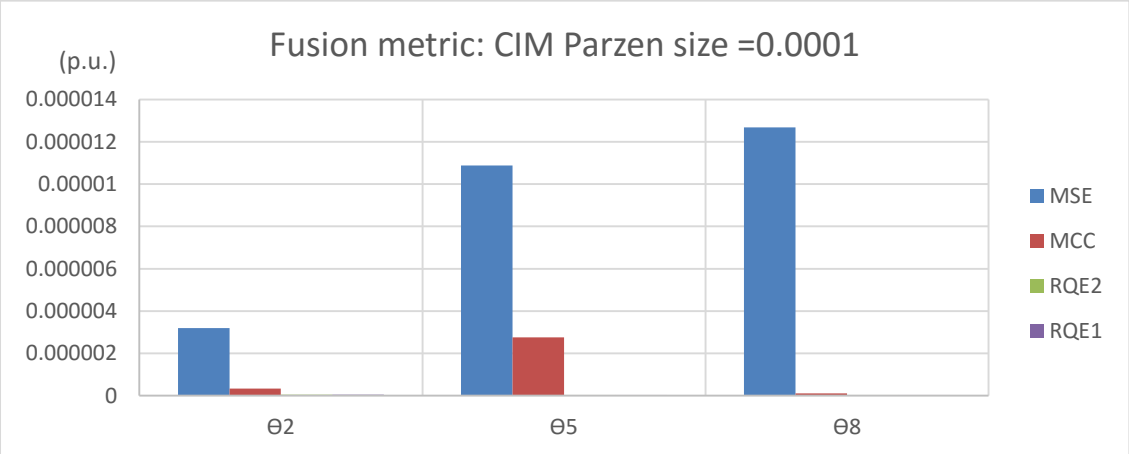

Figure 5-15 - Residual errors in voltage phases using CIM as the fusion metric, with Parzen size 0.0001.

Considering the errors represented in figures 5-6 to 5-13 can be assumed that the Parzen Window's size 1 generate good results, to this study case. Therefore, in the following section, 5.3, the PW size applied to CIM as fusion metric will be 1.

5.2.1 - Results analysis / relevant conclusions.

Comparing the results obtained using the absolute and Euclidean norms as fusion metrics, the most relevant conclusion is the behavior of the Renyi's quadratic entropy (case 2). For most of the components, its residual errors are higher than the errors obtained by the other evaluation metrics. The RQE (case 2) estimation errors are also higher than the noise introduced to the original values (measurements without noise). Still analyzing the fusion metrics L1 and L2, the results obtained by the evaluation metrics MSE and MCC are practically the same, as expected, because there are no outliers and thus both criteria conduct to the same state estimation.

Between MSE and MCC metrics and RQE (case 1) there are no obvious benefits, by the comparison of the results. Even that RQE1 leads to a different state, its residual errors are within the noise introduced to the original values. Thus, it can be considered that the state estimation provided by RQE1 is a reliable state.

As it was already mentioned, using the RQE (case 2), the residual errors reach much higher values than the other evaluation metrics and also higher than the noise introduced to the original values of the measurements (without noise). Also, in RQE case 2, the sensory fusion is not done considering the two groups of measurements separately, in fact, even the evaluation of each system considers interaction with errors from the other group. For that reason, and also considering that the results were worse than the ones obtained by other metrics, the RQE case 2 can be abandoned for these studies.

The CIM as fusion metric is a more delicate situation. Actually, the Parzen Window will interfere with the result, and that PW will consider the deviance between the ideal point and the optimal point found. That distance is unknown and also can have really different ranges to each of its components (variation of fit_1 and fit_2 can assume really different values), so it's risky to use that metric in the fusion step.

5.3 - Sensory fusion optimal point - Measurements with gross errors

In the previous chapter, was already observed the influence of the different evaluation metrics as SE objective function in the presence of gross errors. Now, it was included another step, the fusion step, where the measurements are divided in two groups provided from two different sensory systems. This sub-chapter aims to evaluate the influence the fusion step has in the treatment of the outliers. For this study were used the same measurements as in the previous section. However, now the measurement of the injected power in bus 3 contains a large error, table 5-4. As the measurements were modified also the ideal point was deallocated, in this way, the new ideal point to each evaluation metric can be consulted in table 5-8, considering the PW applied to these metrics with size 2.

Table 5-8 - Ideal point, calculated with σ=2, to the measurements affected with gross error .

	max_{fit1}	max_{fit2}
MSE	-0.682033996077471	-0.165376488179705
MCC	10.915560350232500	5.979774529346460
RQE 1	119.805597386184000	35.489865315248300
RQE 2	184.751933517149000	101.200230478727000

To the fusion metric CIM, the Parzen Window's size was assumed to be 1, because of what was already tested in the previous section. However, also the metrics MCC and RQE, used as evaluation metrics, have the PW concept. Thus, that metrics were tested assuming two values for the PW size, 2 and 0.1. One large window, that allow gross errors to be assumed as part of the measurements, and a thinner window, that ignore the gross errors presented in the measurements.

It's important to highlight that the variation of the Parzen Window in the fusion step (for the CIM) doesn't affect at all the ideal point, once it is not even used to its obtainment. However, varying the PW of the evaluation metrics will change the ideal point, because the fitness values are calculated considering a fixed PW size. Thus, the new ideal points are presented in table 5-9. The MSE doesn't depend on the parameter σ, for that reason it remains the same.

Table 5-9 - Ideal point calculated with σ=0.1 to the measurements affected with gross error .

	max_{fit1}	max_{fit2}
MSE	-0.682033996077471	-0.165376488179705
MCC	9.999992456509830	5.999999141260640
RQE 1	100.999901583335000	35.999983319472600
RQE 2	160.999799009255000	95.999933402185100

In the following figures is possible to see four graphics representing the absolute value of the residual errors. Two of them represent the errors on power injections and lines power flows: one with all the PW sizes and the measurement affected with the gross error, and other ignoring the results for the component affected with a gross error and including only the PW sizes 0.01. The same stands for the graphics of voltage phases.

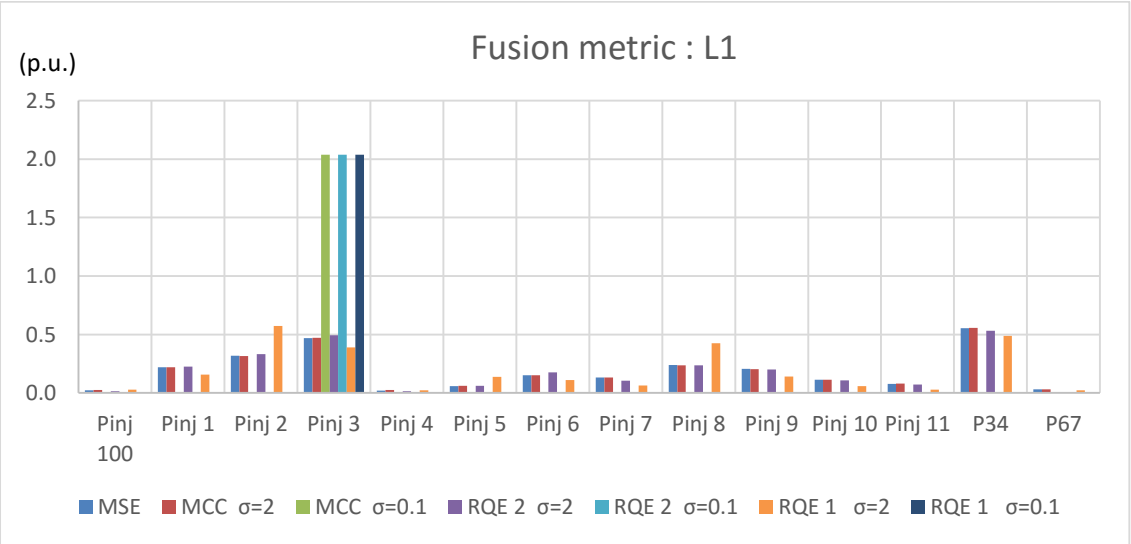

Figure 5-16 - Residual errors in buses power injections and lines power flows using L1 as the fusion metric, with σ= 2 and σ=0.1.

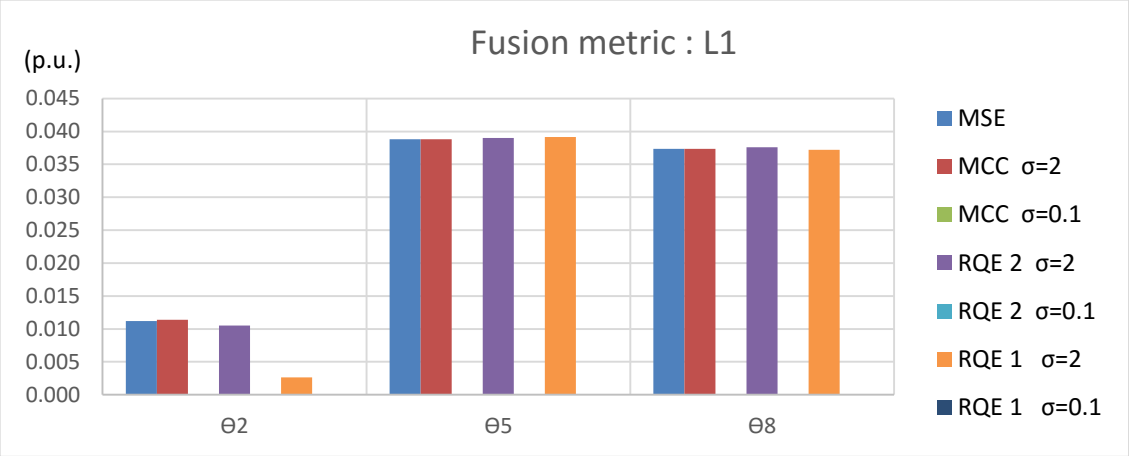

Figure 5-17 - Residual errors in voltage phases using L1 as the fusion metric, with σ=2 and σ=0.1.

It is possible to observe that for MSE, MCC and REQ, with big PW size, all the results get affected by the error in $Pinj_3$. While, for a smaller PW size, 0.1, the MCC and RQE metric ignore that measurement, and only its error is high, all the other component's errors approach to zero.

In the previous image is possible to observe that the decrease of the PW size leads to a better result that MSE, or MCC and RQE with a big PW. However, it's not possible to observe the different effect of each evaluation metric in the errors, due the big error presented in the $Pinj_3$. For that reason, the following image represent only the measurements for PW size 0.1, and ignore the residual error in $Pinj_3$.

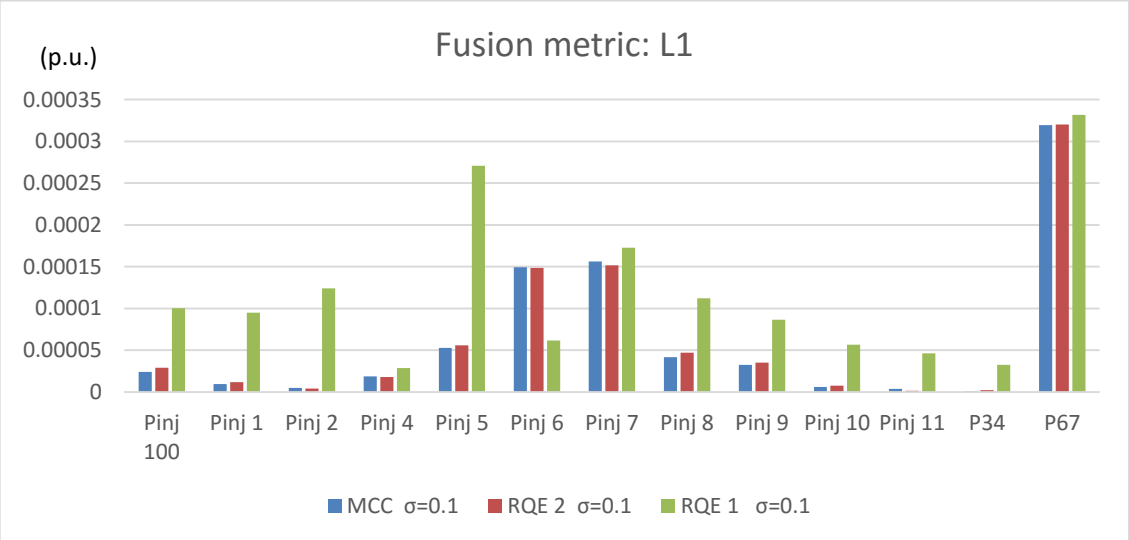

Figure 5-18- Residual errors in buses power injections and lines power flows using L1 as the fusion metric, with σ=0.1, and ignoring the residual error of the component Pinj3.

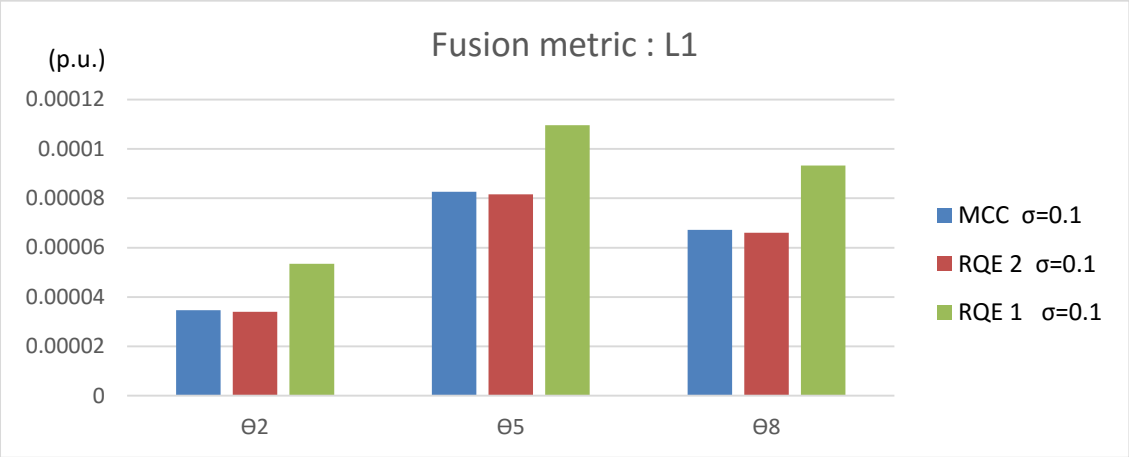

Figure 5-19 - Residual errors in voltage phases using L1 as the fusion metric , with σ=0.1.

Despite what happened in the previous experiment (measurements without gross error), where the RQE case 2 generate residuals far from the ones obtained with MCC and RQE case 1, here, when the PW size decrease the RQE 2 errors approach to same obtained by the MCC. In figures 5-18 and 5-19 the residuals generated by MCC and RQE case 2 are practically the same. The results obtained with the Euclidean norm, L2, as fusion metric are really similar to the ones obtained with the absolute norm, L1. For that reason, the L2 results are in the annex B. The behavior of the fusion metrics L1 and L2 are similar with no obvious benefits in choose one or another metric.

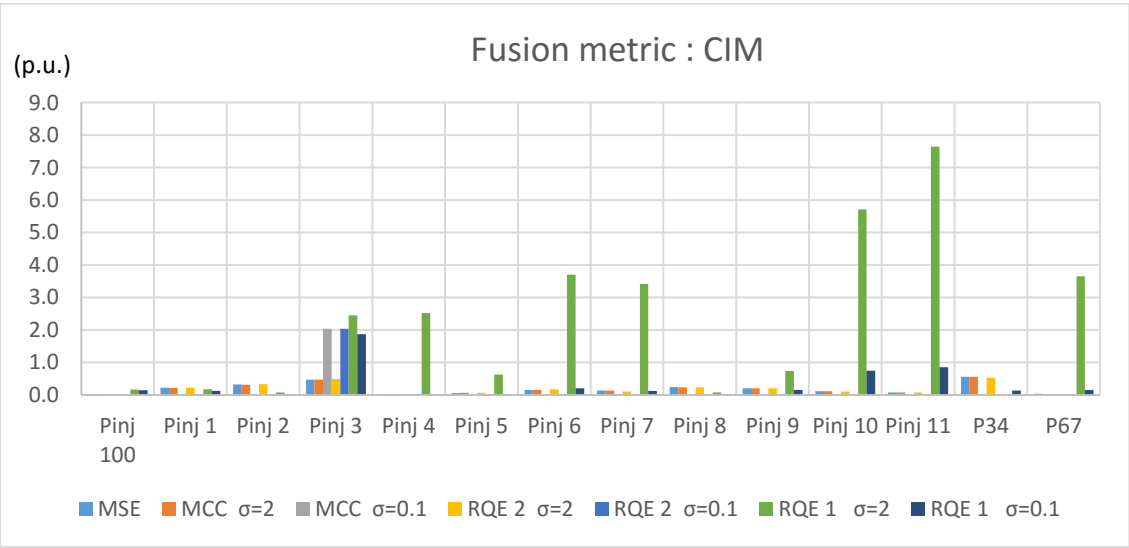

Figure 5-20 - Residual errors in buses power injection and lines power flow, using CIM as the fusion metric, with σ=1. Evaluation metrics with σ=2 and σ=0.1.

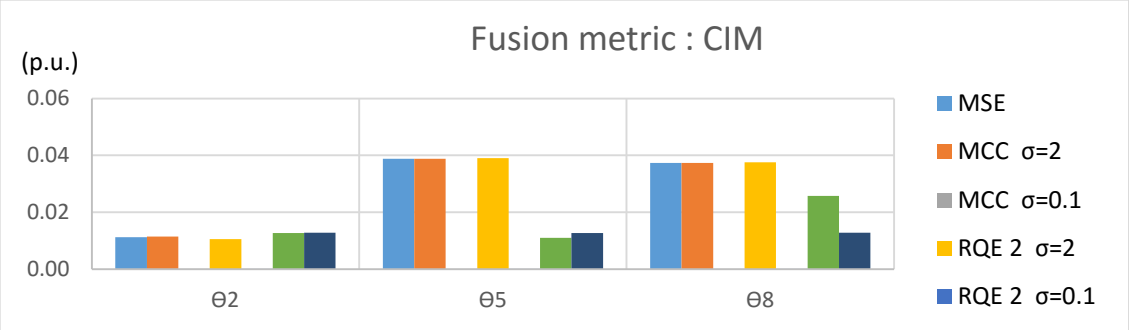

Figure 5-21 - Residual errors in voltage phases, using CIM as the fusion metric, with σ=1. Evaluation metrics with σ=2 and σ=0.1.

Using the CIM as the fusion metric, the residual errors show a different behavior than the ones obtained by L1 and L2 metrics. In particular, the estimations obtained by the evaluation metric RQE (case one) present a big deterioration when applying the CIM as the fusion metric, either for high PW size 2, or small, size 0.1. In annex B is possible to observe the same results presented in figures 5-20 and 5-21, but only for MCC and RQE (case one and two) with σ=0.1, for a better observation of the behavior of RQE (case one) with the fusion metric CIM. In order to obtain better results, different PW size were tested with the evaluation metric RQE (case 1), an improvement of the results was observed for PW size 0.5, figure 5-22 and 5-23. Some other results obtained with a different PW for the RQE (one) can be found in annex B.

Applying a convenient PW size to the evaluation metrics MCC and RQE (case 2), with the fusion metric CIM, the results are reliable and similar to the ones obtained with L1 and L2, figures 5-22 and 5-23. However, the behavior of the evaluation metric RQE (case 1) when combined with fusion metric CIM is not convenient, even for a PW size 0.5.

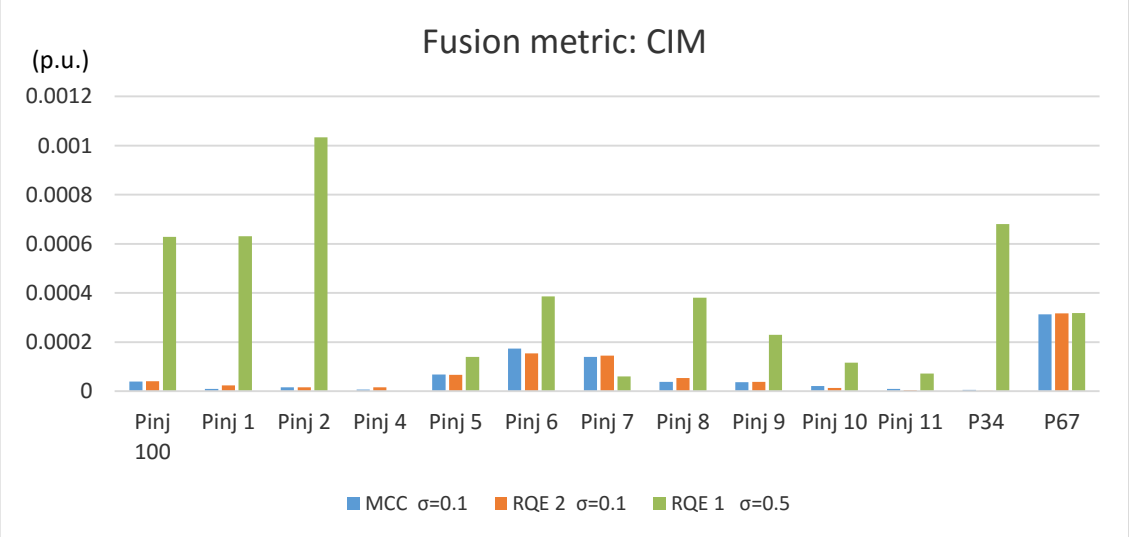

Figure 5-22 - Residual errors in buses power injections and lines power flows, using CIM as the fusion metric (with σ=1) and ignoring the residual error of the component Pinj3. Evaluation metrics MCC and RQE (2) with σ=0.1, RQE (1) with σ=0.5.

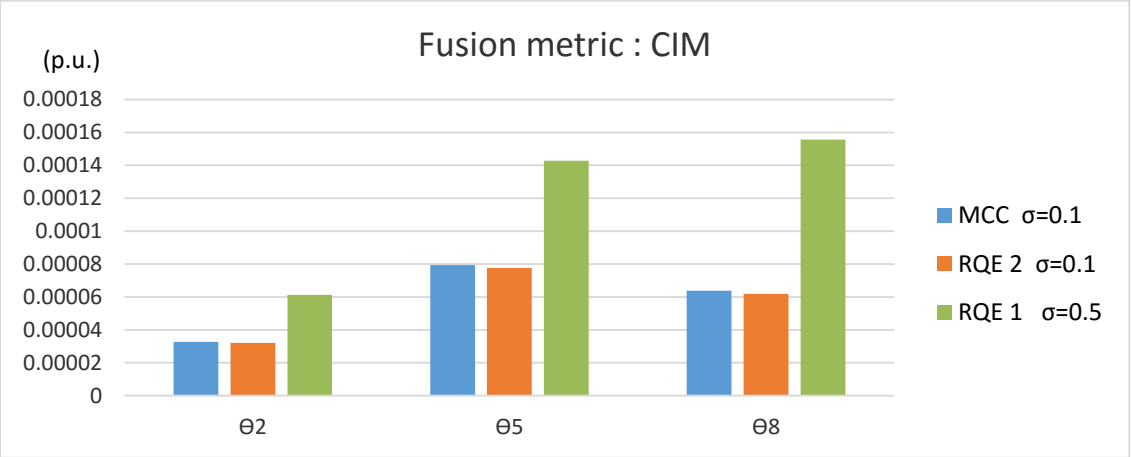

Figure 5-23 - Residual errors in voltage phases, using CIM as the fusion metric (with σ=1). Evaluation metrics MCC and RQE (2) with σ=0.1, RQE (1) with σ=0.5.

From the observation of figures 5-20 to 5-23, it's evident that it's complicated to apply CIM as the fusion metric in the most adequate way, because it is necessary to adjust not only the PW size of the fusion metric but also the PW of the evaluation metrics. Thereby, it is not such easy to identify that parameters as it is when using the absolute norm (L1) or Euclidean norm (L2).

The residual errors are the most important thing to guarantee with the state estimation. However, the iterations or the duration of the process is also an important parameter, taking in account that the problem has been solved by a meta-heuristic method- EPSO [30]. Thus, in tables 5-10 and 5-11 can be found the amount of fitness evaluation each metrics combination took until reach the optimal point, and the time all that process took.

Table 5-10 - Number of fitness evaluation each combination took to find the optimal point.

fitness evaluations	MSE	MCC (σ=2)	RQE1 (σ=2)	RQE2 (σ=2)	MCC (σ=0.1)	RQE1 (σ=0.1)	RQE2 (σ=0.1)
L1	4.82E+06	2.70E+06	2.16E+06	1.78E+06	1.46E+06	2.98E+06	2.31E+06
L2	7.65E+06	6.46E+06	1.53E+07	1.31E+07	8.61E+05	2.13E+06	3.25E+06
CIM (σ=1)	1.85E+06	2.49E+06	9.41E+04	3.46E+06	2.41E+06	3.19E+05	2.89E+06

Table 5-11 - Duration of the process each combination took to find the optimal point.

Duration (s)	MSE	MCC (σ=2)	RQE1 (σ=2)	RQE2 (σ=2)	MCC (σ=0.1)	RQE1 (σ=0.1)	RQE2 (σ=0.1)
L1	831	505	422	364	232	494	408
L2	1370	1002	2516	2347	134	354	564
CIM (σ=1)	295	389	16	596	376	54	504

The evaluation metric RQE (case 1), with PW size 0.1 and combined with the fusion metric CIM, doesn't lead to a good result, so it's irrelevant that it took just few fitness evaluations and less time than the others.

The absolute norm, L1, took most of the times less time to converge than the Euclidean metric, L2, and that's not surprising since the expression is simpler. However, the results are not concluding in which norm will always perform a better time. Also, it is important to mention that, for now, the problem has been solved by EPSO because the purpose of this thesis is to

propose new perspectives about the sensory fusion. For that reason, a meta-heuristic method is enough to solve the proposed problem, just to show that the new methods and metrics are applicable, convenient and to show them advantages. The need for a further work capable to solve the problem by a differential method, consuming much less time is evident.

5.3.1 - Results analysis / relevant conclusions.

The application of the developed fusion method to the SE process doesn't interfere with the capability of criteria as minimum Renyi's quadratic error or maximum Correntropy criterion ignore outliers.

From the tested fusion metrics, the CIM is the most complicated to manage, because is necessary to assign values to the PW size of the fusion metric and also of the evaluation metric. As it was observed previously, it's not so simple to identify the right PW size as it is when using the absolute and the Euclidean norms, L1 and L2, respectively. Besides that, the CIM doesn't show any advantage on its application. Moreover, even without gross errors it's complicated to find a PW size that shows totally convenience to both fitness evaluation functions (fitness evaluation of group one and two). In fact, the highness of the distance between the optimal and the ideal point is unknown, so doesn't make much sense to limit it through a PW method.

With the introduction of an outlier, the evaluation metric RQE (case 2) generated errors close to the ones obtained with MCC. Despite the good results obtained in this section, the RQE2 meaning is not what has been sought in this thesis. The idea of the fusion step is to fusion information descendant from both sensory systems separately. RQE (case 2) promotes the interaction of each component's error with all the errors from its own group and also with the other one. So it's not separately. Moreover, it doesn't present any advantage in comparison with the other metrics, actually, in the previous section it presented results even worse than the other criteria. For that reasons this approach to the RQE will be abandoned in the next chapter.

5.4 - Comparison between simple fusion and the developed fusion method

As it was already mentioned, the inclusion of PMU in the monitoring system can be done in different ways. One possibility is the one that has been discussed on the previous sections of this chapter. Another way, the simplest and most common, is to include the measurements obtained by PMU in the same vector of SCADA measurements, computing the state estimation with all the measurements together, without any distinction based on its origin, that is, the proceed in chapter 3.

This section aims to compare the results obtained by both sensory fusion methods. For that, the results of chapter 3 are compared with the result of the first section of this chapter, for the case with no gross errors. The objective here is not to compare the effect of the different metrics in the presence of gross errors, that was already explored in chapters 3.5 and 5.1.2., but to compare the two different approaches of sensory fusion. In the following figures is possible to observe the results for the simple fusion (without any distinguish by the measurements' origin) and the developed fusion method in this chapter, with fusion metrics L1 and L2. This three cases were compared under the different evaluation metrics: MSE, MCC and RQE, with a PW size 0.1.

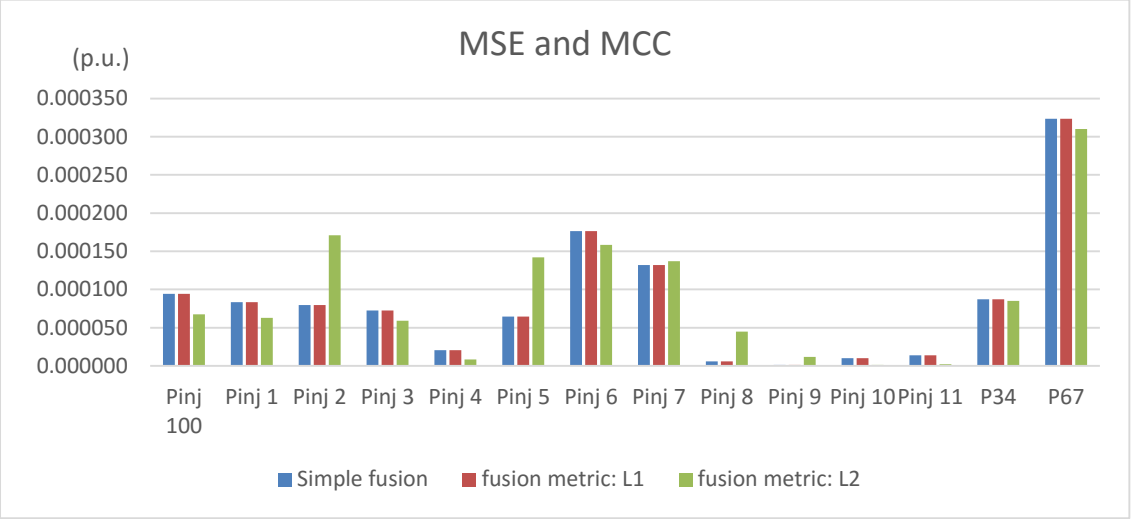

Figure 5-24 - Power injections residual errors obtained with different fusion methods. MCC, with σ=0.1, and MSE as evaluation metrics.

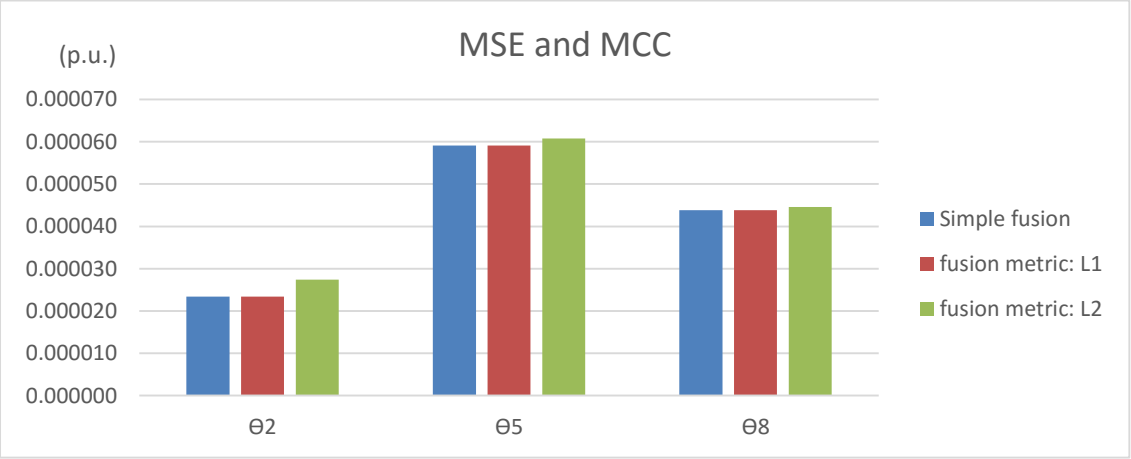

Figure 5-25 - Voltage phases residual errors obtained with different fusion methods. MCC, with σ=0.1, and MSE as evaluation metrics.

As was already seen, when there are no gross errors, the results obtained by MCC or minimum MSE are the same, for that reason the previous figures, 5-24 and 5-25, contains that results in simultaneous, so there's no replied graphics, once the ones obtained by MSE and MCC are the equal.

Using the MCC or MSE, the L2 norm perform different residuals than the L1 and the simple fusion. As already analyzed, the results obtained by the L2 norm are acceptable and reliable.

The important thing to mention here is that the absolute norm is preforming exactly the same results as the simple fusion, for the evaluation metrics MSE and MCC. If we look to those expressions again, table 5-1, we can easily conclude that the minimization of the distance between the optimal and the ideal point, by the L1 norm, can be easily simplified by the expressions in table 5-12.

Table 5-12 – Simplified objective function using the metric L1 to find the closest point to the ideal one.

Objective function		**Fusion metric: L1**
Evaluation metric	MSE	min $\sum_{i\epsilon n1}(z_i-\hat{z}_i)^2+\sum_{i\epsilon n2}(z_i-\hat{z}_i)^2$
	MCC	max $\sum_{i\epsilon n1} e^{-\frac{(z_i-\hat{z}_i)^2}{2\sigma^2}}+\sum_{i\epsilon n2} e^{-\frac{(z_i-\hat{z}_i)^2}{2\sigma^2}}$
	RQE1	max $\sum_{i\epsilon n1}\sum_{j\epsilon n1} e^{-\frac{(r_i-r_j)^2}{2\sigma^2}}+\sum_{i\epsilon n2}\sum_{j\epsilon n2} e^{-\frac{(r_i-r_j)^2}{2\sigma^2}}$

In the previous table the parameter σ refers to the PW size and n_a is the set of measurements provided by the sensory system a. N consists in the aggregation of both measurement systems, 1 and 2. $N=n_1 U n_2$.

In fact, for MSE and MCC the developed fusion method, using the L1 norm, presents the same mathematical approach as the simple fusion. The main advantage of the developed method is to easily convert the L1 norm into a stochastic fusion, were parameters α and ß are the probabilities of each sensory system to conduct to the true state, computing a state estimation with minimum error.

The same equality between the simple fusion and the developed fusion method is not true when the evaluation metric is the RQE. In fact, in the simple fusion each component error interacts with all the other component errors, independently of the measurement origin. In the developed fusion method, each error only interacts with errors from the same sensory system.

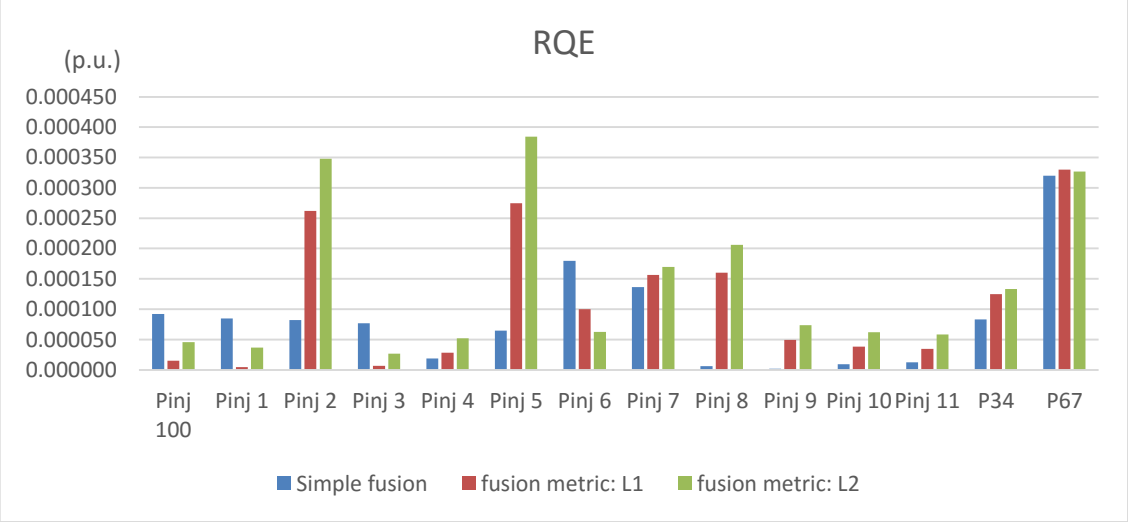

Figure 5-26 - Power injections residual errors obtained with different fusion methods with RQE as evaluation metric, with σ=0.1.

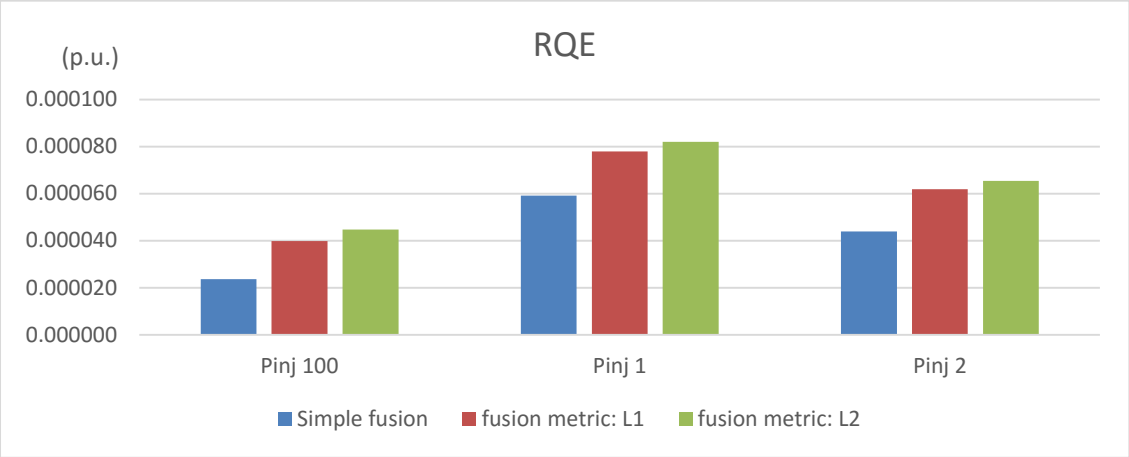

Figure 5-27 – Voltage phases residual errors obtained with different fusion methods with RQE as evaluation metric, with σ=0.1.

As it can be seen by the previous graphics, the developed fusion method generates a different state than the one obtained with the simple fusion, when the evaluation metric is RQE. Despite the difference, both points are acceptable and reliable once the residuals are within the noise introduced to the original values.

To decide which fusion metric apply in the next chapter two studies were made. The errors' entropy and the mean absolute percentage error were calculated to compare the results obtained with both sensory methods, tables 5-13 and 5-14. The errors' entropy was calculated based on the Renyi's quadratic error, already presented in Chapter 3.1 -, applying a PW with σ=0.1. The mean absolute percentage error was calculated by

$$MAPE = \frac{1}{n}\sum_{k=1}^{n}\left|\frac{z_k - \hat{z}_k}{z_k}\right|, \tag{5.1}$$

where Z_k is the true value of the component k, and $\widehat{Z_k}$ is its estimated value.

Table 5-13- Errors' entropy and mean absolute error for the results obtained with simple fusion.

Fusion	**Simple fusion**		
	MSE	MCC	RQE
Renyi's quadratic entropy	-1.3836459676	-1.3836459676	-1.3836459681
Mean Absolute error	0.40%	0.40%	0.39%

Table 5-14 - Errors' entropy and mean absolute percentage error for the results obtained with fusion metrics L1 and L2.

Fusion	**absolute norm**			**Euclidean norm**		
	MSE	MCC	RQE	MSE	MCC	RQE
RQE	-1.38364597	-1.38364597	-1.38364560	-1.38364595	-1.38364595	-1.38364525
MAPE	0.40%	0.40%	0.77%	0.47%	0.47%	1.08%

It's clear that the objective is to decrease the mean error and the information contained in the error, that is, to decrease the entropy. In this perspective, the absolute norm shows a better perform when compared with the Euclidean norm.

5.5 - Chapter conclusions

In this chapter was possible to observe the influence of different fusion metrics on the behavior of the diverse evaluation metrics. The first conclusion is that the fusion metric CIM doesn't bring any advantage. In fact, it brings the disadvantage of interfering with PW size of the evaluation metrics. Moreover, the fitness evaluation functions of both sensory systems have different ranges, being difficult to select a convenient PW size to the fusion metric CIM.

When comparing the absolute and Euclidean norms, it can be said that both fusion metrics generate reliable and acceptable results. However, L1 leads to scenarios with less entropy and a lower mean absolute error. The absolute norm, L1, has also the convenience of being easily simplified by the expressions presented in table 5-12. Thus, it's possible to compute the estimation without the need of previously determine the ideal point, making the process simpler and faster.

The absolute norm, when simplified, shows the same mathematical expression as the simple fusion, when the evaluation metrics are MCC and MSE. Despite the equal expression, the fusion method shows the advantage of being easily converted into a stochastic fusion, as already explained and proposed in the previous chapter. For that stochastic fusion a many research and studies are required. However, this new perspective, here developed, allows a step into that path.

Chapter 6

Application of the fusion method to the AC model

Until now, in the previous chapters, all the studies were made considering a non-resistive network. As the network used had zero resistance, the DC model presented a good approach. Therefore, the DC model was applied to the developed sensory fusion method. In order to evaluate the developed method performance, and make some decisions, some experiences were provided: comparison between all the evaluation metrics, trust variation among the different sensory systems and the selection of the optimal fusion point.

Now, for more reliable conclusions, the developed fusion method must be tested in a real network. In this chapter, the developed fusion method is tested, comparing different evaluation metrics, in a network with high resistance. The network utilized in this chapter has the same topology as the one presented in chapter 3 (used until now); however, this network presents a high resistive component. The lines parameters of the actual network can be found in annexes A.1. The distributed generation and loads profiles are presented in the same annex, in section 2, and the measurements in section 3.1. The mentioned network is adapted from the "Benchmark Systems for Network Integration of Renewable and Distributed Energy resources", from Cigre [29]. It presents typical parameters of a European medium voltage distribution network. Assuming a real network, the DC model is not a good approach, hence the AC model is applied to the state estimations provided in this chapter.

Once the main goal of this thesis is to provide a proof of concept of the developed method, the state estimation problem has been solved by a meta-heuristic algorithm-EPSO [30]. A further work can develop an alternative resolution to the proposed method. In order to make the process faster and applicable to real time applications, alternative resolutions can integrate derivatives methods as Gauss-Newton. However, the purpose of this thesis is to show that's the proposed method is applicable, and to show its advantages. For that purpose, the EPSO algorithm is more than enough.

6.1 - Measurements

In annex A.3 is possible to see how the measurements, used in this chapter, were obtained. The measurements provided to integrate the AC model, for each one of the tests, can be consulted in the following table 6-1.

Table 6-1- Measurements used in the AC model without gross errors.

	Bus	Power Flow result	Measurement affected with noise (without PMU)
P_{inj} (p.u.)	100	1.121140	1.122810
	1	0.041322	0.039534
	2	-0.807506	-0.809074
	3	-0.569125	-0.572176
	4	-0.048778	-0.045104
	5	-0.048648	-0.043448
	6	1.449080	1.451090
	7	-0.291667	-0.291622
	8	-0.059664	-0.062184
	9	-0.430091	-0.429711
	10	-0.315175	-0.318894
	11	-0.032993	-0.033324
Q_{inj} (p.u.)	100	-0.238982	-0.236943
	1	0.455981	0.457423
	2	-0.265455	-0.267485
	3	-0.182535	-0.182715
	4	-0.010337	-0.011612
	5	-0.009668	-0.009689
	6	0.590705	0.588743
	7	-0.096752	-0.094147
	8	-0.011773	-0.010499
	9	-0.109040	-0.109329
	10	-0.096427	-0.098370
	11	-0.006464	-0.006126
ϴ (rad)	2	-0.012750	-
	5	-0.014990	-
	8	-0.015320	-
V (p.u.)	2	0.997079	0.998354
	5	0.996383	0.999383
	8	0.995781	0.992141
P_{ij} (p.u.)	3 4	0.169774	0.166874
	6 7	-0.886705	-0.885805
Q_{ij} (p.u.)	3 4	0.158326	0.159126
	6 7	-0.344520	-0.349320
I_{inj}(real) (p.u.)	2	-0.806411	-
	5	-0.048673	-
	8	-0.059729	-
I_{inj}(im) (p.u.)	2	-0.276537	-
	5	-0.010434	-
	8	-0.012739	-

The measurement sets that contain gross errors are presented in the table 6-2. These sets will be used to test the behavior of the evaluation metrics in the presence of outliers, in a real based network.

Table 6-2 - Measurements used in the AC model with gross errors.

	Bus	Measurements (missing measurement) (with PMU)	Measurement (Power flow inversion) (with PMU)
P_{inj} (p.u.)	100	1.122810	1.122810
	1	0.039534	0.039534
	3	-0.572176	-0.572176
	4	-0.045104	-0.045104
	6	0.000000	1.451090
	7	-0.291622	-0.291622
	9	-0.429711	-0.429711
	10	-0.318894	-0.318894
	11	-0.033324	-0.033324
Q_{inj} (p.u.)	100	-0.236943	-0.236943
	1	0.457423	0.457423
	3	-0.182715	-0.182715
	4	-0.011612	-0.011612
	6	0.588743	0.588743
	7	-0.094147	-0.094147
	9	-0.109329	-0.109329
	10	-0.098370	-0.098370
	11	-0.006126	-0.006126
ϴ (rad)	2	-0.012826	-0.012826
	5	-0.015083	-0.015083
	8	-0.014795	-0.014795
V (p.u.)	2	0.998354	0.998354
	5	0.999383	0.999383
	8	0.992141	0.992141
P_{ij} (p.u.)	3 4	0.166874	0.166874
	6 7	-0.885805	0.885805
Q_{ij} (p.u.)	3 4	0.159126	0.159126
	6 7	-0.349320	0.349320
I_{inj}(real) (p.u.)	2	-0.806061	-0.806061
	5	-0.048913	-0.048913
	8	-0.059329	-0.059329
I_{inj}(im) (p.u.)	2	-0.276589	-0.276589
	5	-0.010506	-0.010506
	8	-0.012667	-0.012667

6.2 - The inclusion of the PMU measurements in state estimation

PMU has been theme of several discussions, many research about its integrations in the network's monitoring system have been conducted [32].

Until now, this work has been developing a fusion method capable to integrate PMU measurements in the measurement system. However, it's important to prove that the inclusion

of that measurements are really favorable. Thus, a first case was tested, where all the buses have power injection sensors, table 6-1-measurements without PMU. Then a second case was studied, where the injected power and voltage magnitude sensors on buses 2, 5 and 8 where replaced by PMU's. In this way, measurements of these buses were replaced by injected currents and voltages (magnitude and phases) measurements, table 6-1-Measurements with PMU.

The residual errors obtained in each case, without and with PMU's, is represented in the following figures 6-1 to 6-5. These results, when the evaluation metrics were RQE or Correntropy, were obtained for a PW with size 1.

As it was already seen in previous chapters, when there are no outliers, MCC and MSE generate the same results. In following graphics, the MSE is not represented just for a matter of simplification of the results, once the state estimation obtained by the minimization of MSE or maximum Correntropy were practically equal.

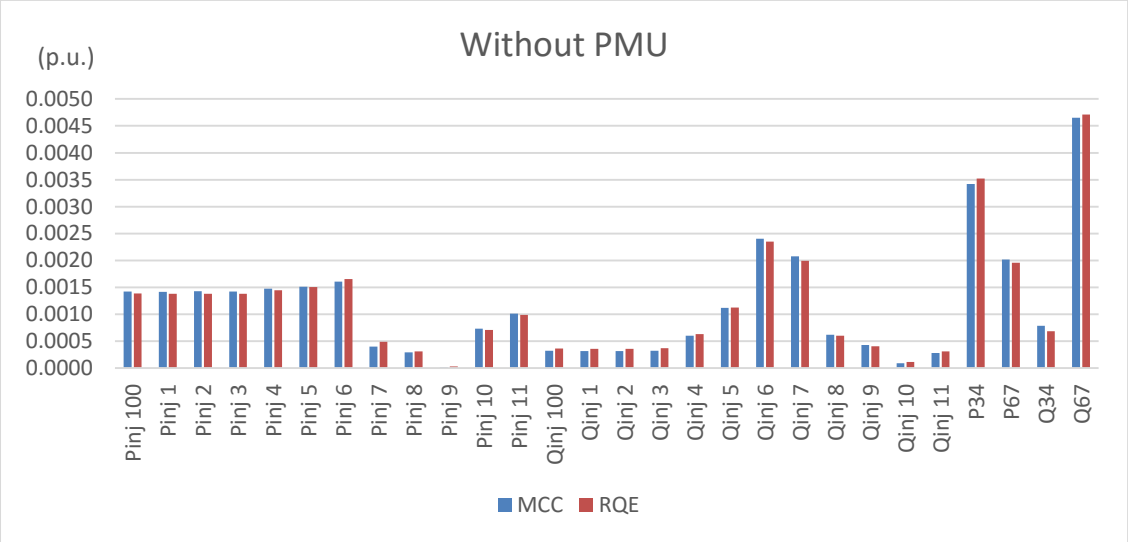

Figure 6-1 - Absolute residual errors in power injection, obtained without PMU.

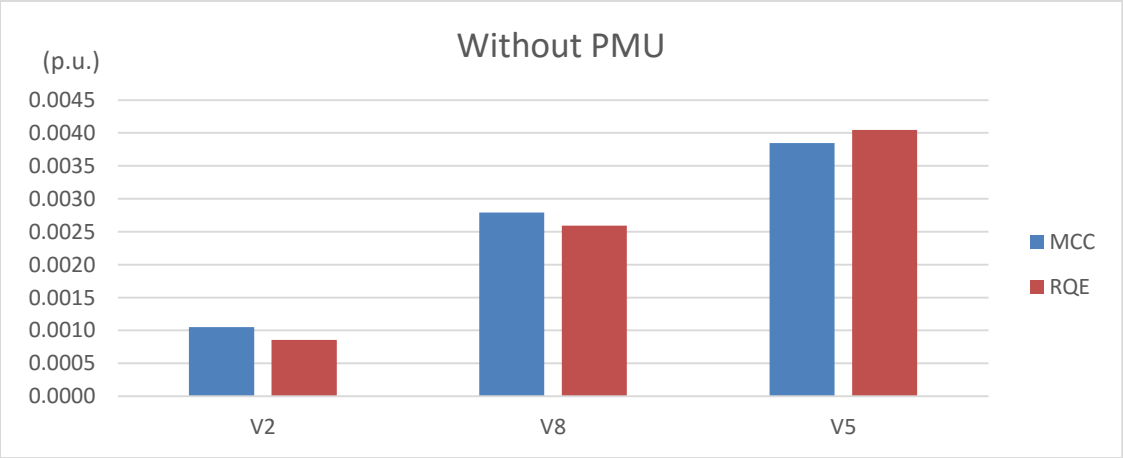

Figure 6-2 - Absolute residual errors in voltage magnitude, obtained without PMU.

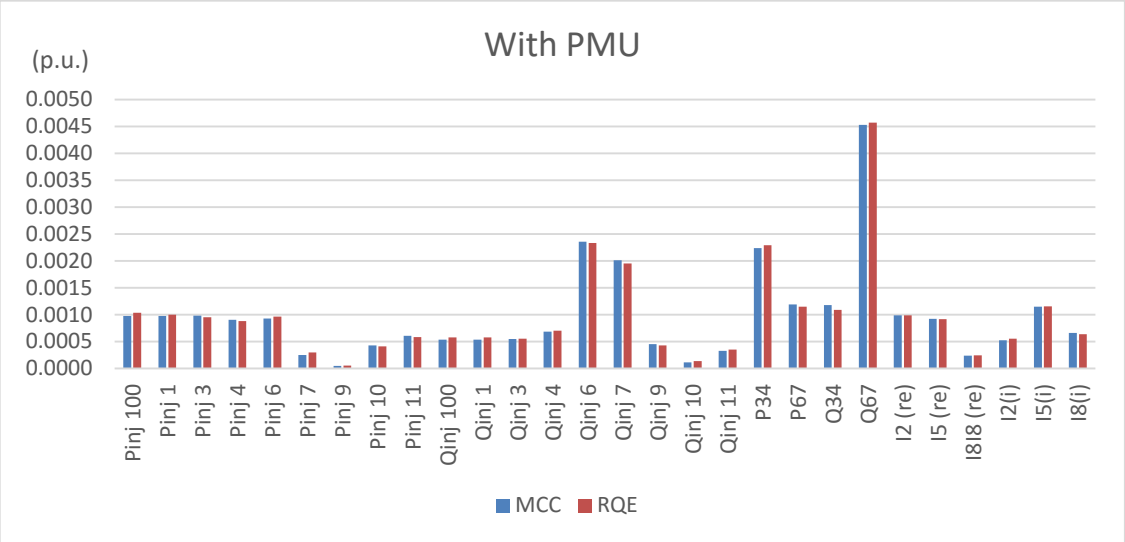

Figure 6-3- Absolute residual errors in power and current injections and lines power flow, obtained with PMU.

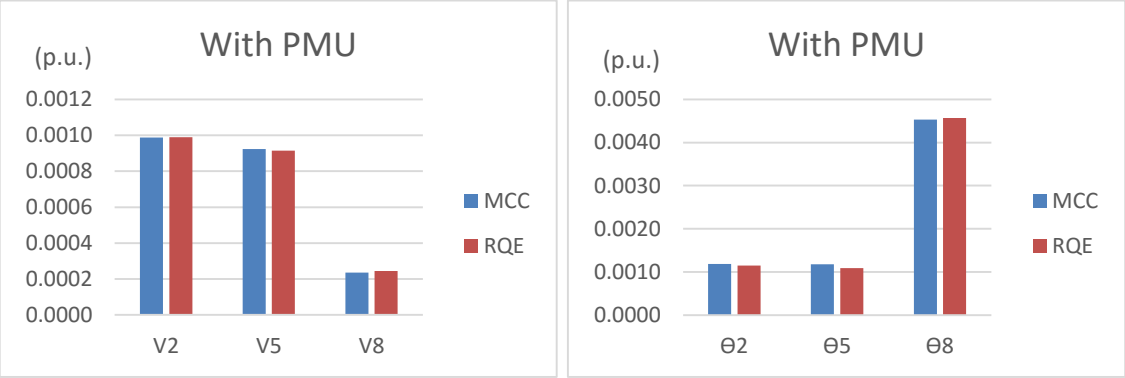

Figure 6-4 – Absolute residual errors in voltage magnitudes, on the left, and on voltage phases, on the right. Errors obtained with PMU.

From the analysis of the residual errors represented on the previous graphics, it can be said that, in a general mode, the errors have decreased with the introduction of PMU; However, the measured components are not the same in both cases. For instance, the currents injections only appear with the introduction of the PMU, and the power injection, in buses 2,5 and 8, were only measured when PMU where not connected to that buses. Thus, to analyze the PMU integration effect on the results, it's necessary to observe not only the residual errors (difference between estimations and measurements) but also the true estimation errors. In this case is possible to observe the real error, once the measurements were generated from real power flow values, presented in annex A.3.1.

It's important to mention that the estimations were computed based on measurements affected with noise. The state variables, voltage phases and magnitude, obtained in each case, with and without PMU, were used to calculate estimates to all the components present in both cases (figures 6-1 to 6-5). In this case, the estimated values were not compared with the measurements (that are already affected with errors) but with the original values. The error between those state estimations and the original values can be observed in figures 6-6 to 6-9.

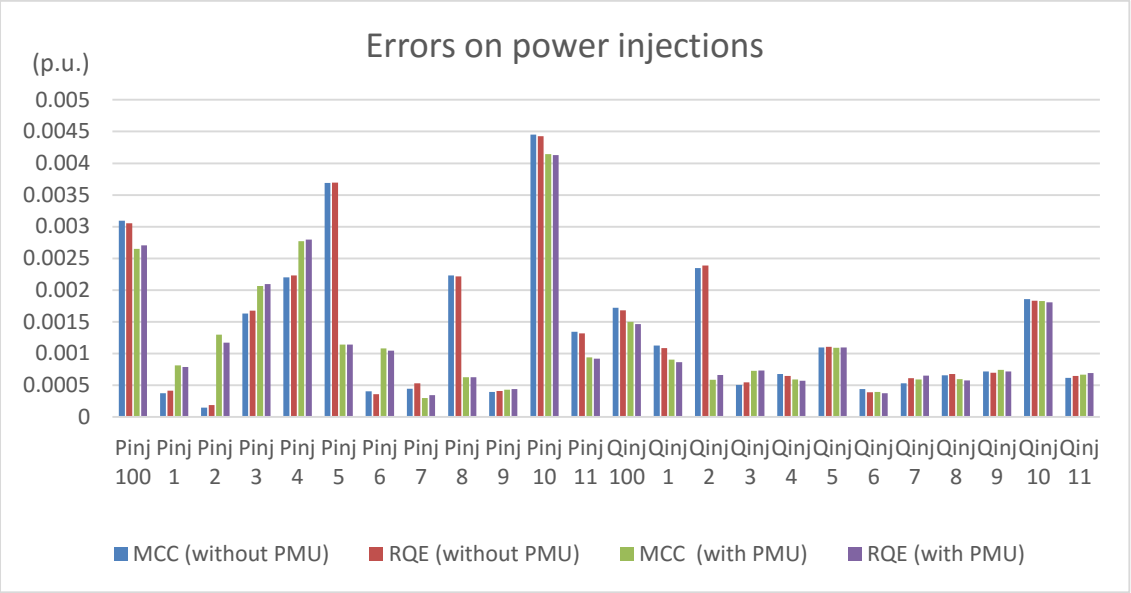

Figure 6-5 - Errors between estimations and original values of injected power.

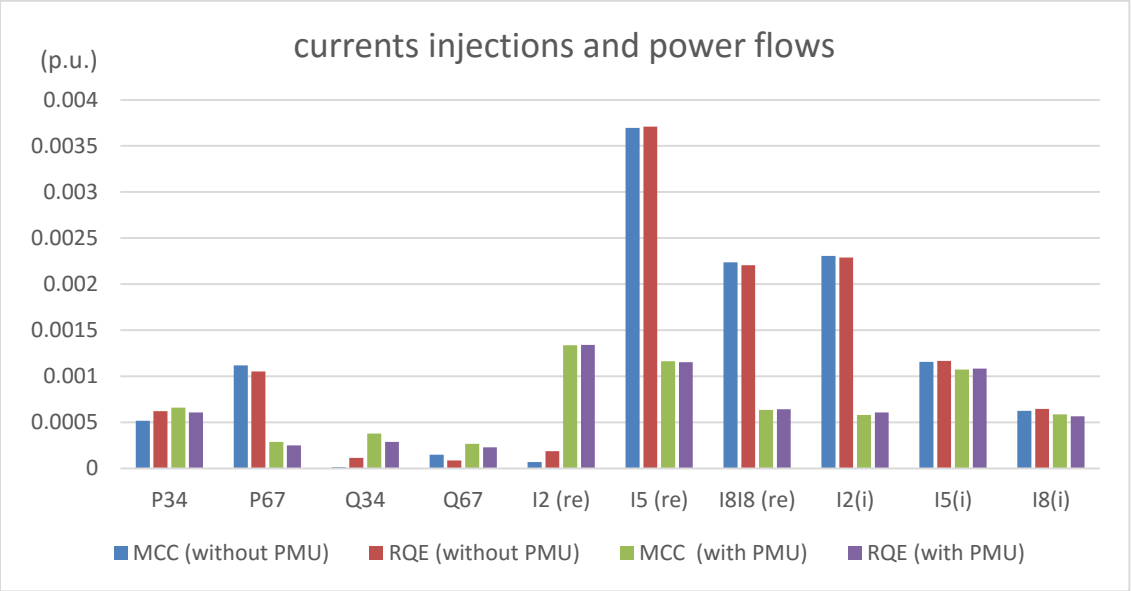

Figure 6-6 - Errors between estimations and original values of injected currents and lines power flows.

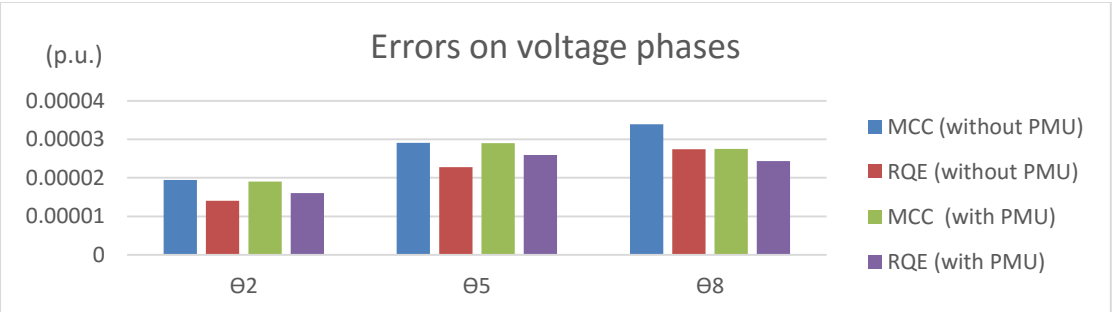

Figure 6-7 - Errors between estimations and original values of voltages phases.

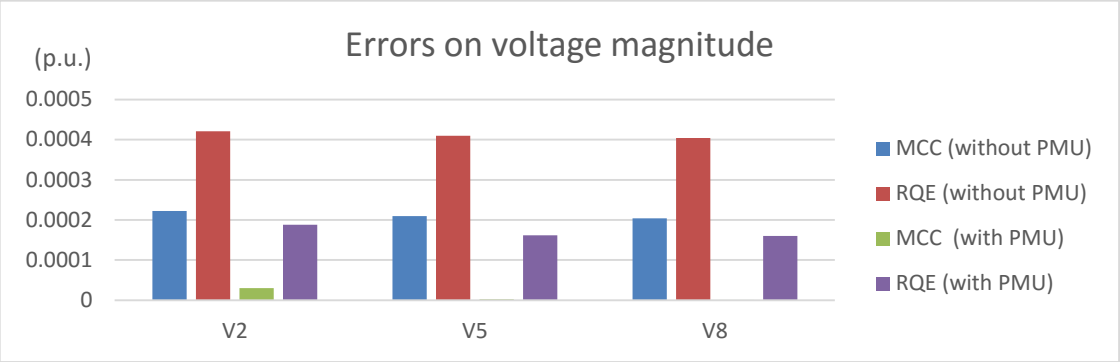

Figure 6-8 - Errors between estimations and original values of voltages magnitudes.

From the observation of the figures 6-5 to 6-7, it's easy to conclude that the replacement of conventional sensors by PMU brigs better results, with lower errors and more reliable state estimations.

For a better general perspective of the errors, obtained by the different sensor systems, the mean absolute percentage error (MAPE) was calculated to each case. Once the results obtained by the minimization of MSE were really similar to the ones obtained by MCC, in the previous graphics, the representations MSE errors were ignored, in order to save some space. In table 6-2 is possible to see the MAPE to each sensor system, obtained using different SE criteria. Once both criteria lead to the same solution, as expected, the MAPE obtained for the MSE and MCC criterions were the same.

The expression used to calculate the MAPE was

$$MAPE = \frac{1}{n}\sum_{k=1}^{n}\left|\frac{z_k - \hat{z}_k}{z_k}\right|, \tag{6.1}$$

where Z_k is the true value of the component k, and $\widehat{Z_k}$ is its estimated value. Also the RQE was calculated for the diverse results, by the expression already presented in chapter 3.

Table 6-3 - Mean percentage error and entropy calculated to each sensor system

	Without PMU			**With PMU**		
	MSE	MCC	RQE	MSE	MCC	RQE
MAPE	2.27%	2.27%	2.29%	1.83%	1.83%	1.83%
RQE	-3.09472	-3.09472	-3.09472	-3.09483	-3.09483	-3.09483

If by the observation of the previous graphics could leave some doubt, about the advantage or disadvantage of PMU inclusion in measure system, the data in table 6-2 clarify it. The inclusion of PMU in the sensor system decreases the error of the estimations, for both evaluation metrics, MCC and RQE, and it also decreases de error's entropy. Thus, one must say that, PMU increases the reliability of the measurement system.

6.3 - State estimation using the fusion method

In chapter 4 were proposed a fusion method with capability to assigning stochastic values to each sensory system, according to the probability of each system to conduct to the real state. As the information about those stochastic values is not yet available, in chapter 5, was

proposed some fusion metrics to find an optimal fusion point. This point is over the Pareto-front resulted from the trust variation on each sensory system. That optimal point can be found by minimizing the absolute distance between the ideal and the optimal point we are searching for. For this case, the SE objective functions, depending on different evaluation metrics, are the ones presented in the following table.

Table 6-4 - Objective function to find the optimal point by the L1 metric.

Objective function		**Fusion metric: L1**
Evaluation metric	MSE	$\min \left\lvert max_{fit1} - \sum_{i\epsilon n1}(z_i - \hat{z}_i)^2 \right\rvert + \left\lvert max_{fit2} - \sum_{i\epsilon n2}(z_i - \hat{z}_i)^2 \right\rvert$
	MCC	$\min \left\lvert max_{fit1} - \sum_{i\epsilon n1} e^{-\frac{(z_i - \hat{z}_i)^2}{2\sigma^2}} \right\rvert + \left\lvert max_{fit2} - \sum_{i\epsilon n2} e^{-\frac{(z_i - \hat{z}_i)^2}{2\sigma^2}} \right\rvert$
	RQE	$\min \left\lvert max_{fit1} - \sum_{i\epsilon n1}\sum_{j\epsilon n1} e^{-\frac{(r_i - r_j)^2}{2\sigma^2}} \right\rvert + \left\lvert max_{fit2} - \sum_{i\epsilon n2}\sum_{j\epsilon n2} e^{-\frac{(r_i - r_j)^2}{2\sigma^2}} \right\rvert$

As it was seen in chapter 5, those expressions can be simplified by the ones presented in table 6-5. In both tables the parameter σ refers to the PW size and n_a is the set of measurements provided by the sensory system a. N consists in the aggregation of both measurement system, 1 and 2. $N=n_1 U n_2$.

Table 6-5 - Simplification for the SE objective function, given by the L1 metric.

Objective function		**Fusion metric: L1**
Evaluation metric	MSE	$\min \sum_{i\epsilon n1}(z_i - \hat{z}_i)^2 + \sum_{i\epsilon n2}(z_i - \hat{z}_i)^2$
	MCC	$\max \sum_{i\epsilon n1} e^{-\frac{(z_i - \hat{z}_i)^2}{2\sigma^2}} + \sum_{i\epsilon n2} e^{-\frac{(z_i - \hat{z}_i)^2}{2\sigma^2}}$
	RQE	$\max \sum_{i\epsilon n1}\sum_{j\epsilon n1} e^{-\frac{(r_i - r_j)^2}{2\sigma^2}} + \sum_{i\epsilon n2}\sum_{j\epsilon n2} e^{-\frac{(r_i - r_j)^2}{2\sigma^2}}$

When using the evaluation metrics MSE and MCC, the objective functions are mathematically equivalent to the simple fusion. For the RQE, the simplification doesn't lead to the same mathematical expression as the simple fusion, because of the following reason: in the simple fusion, the RQE promotes the interaction between all the errors, independent to the sensory system each error belongs, while, in the developed fusion method, the sensory systems are separated. That is, for RQE, each error interacts only with the errors within the same sensory system. The main advantage, of the fusion method compared to the simple fusion, is its ability to easily convert into a stochastic method, by assigning weights to each one of the sensory systems term. In order to find proper ways to calculate those stochastic values, new research and studies must be performed.

In this way, the developed fusion method was tested with the measurements presented in table 6-1- measurements with PMU. The residual errors obtained with each evaluation metric can be found in the following graphics.

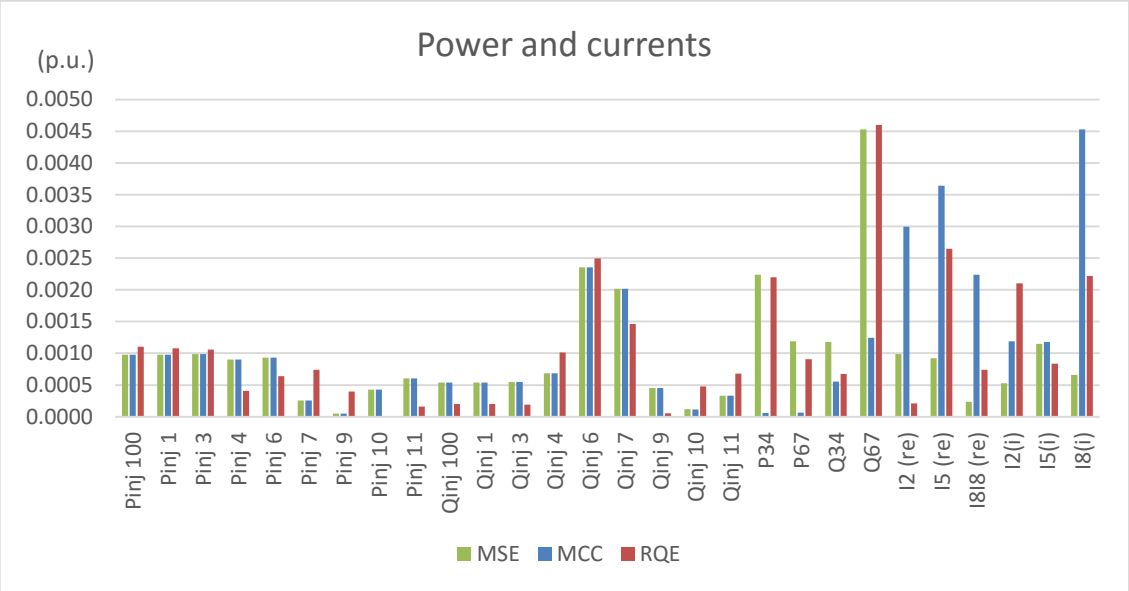

Figure 6-9 - Absolute value of residual errors in injected power and currents and lines power flow for the AC model, without gross errors.

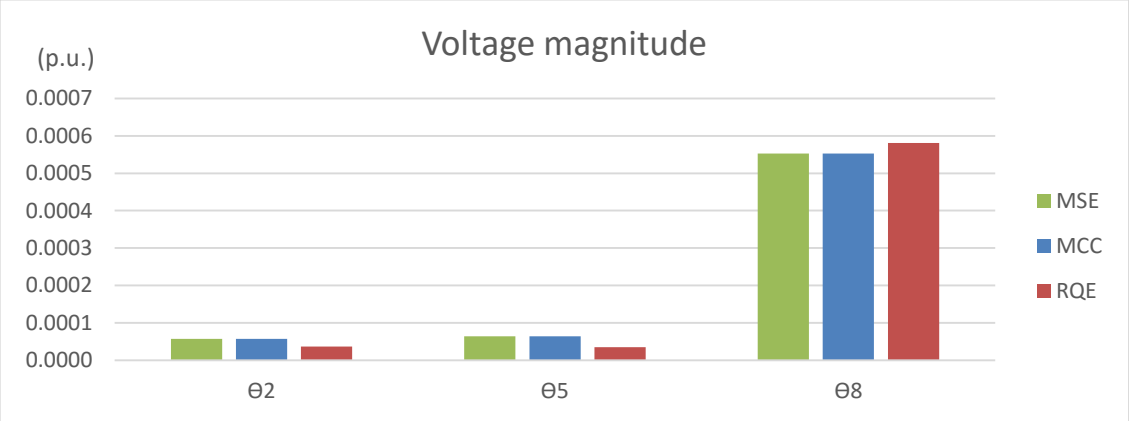

Figure 6-10- Absolute value of residual errors in voltage phases for the AC model, without gross errors.

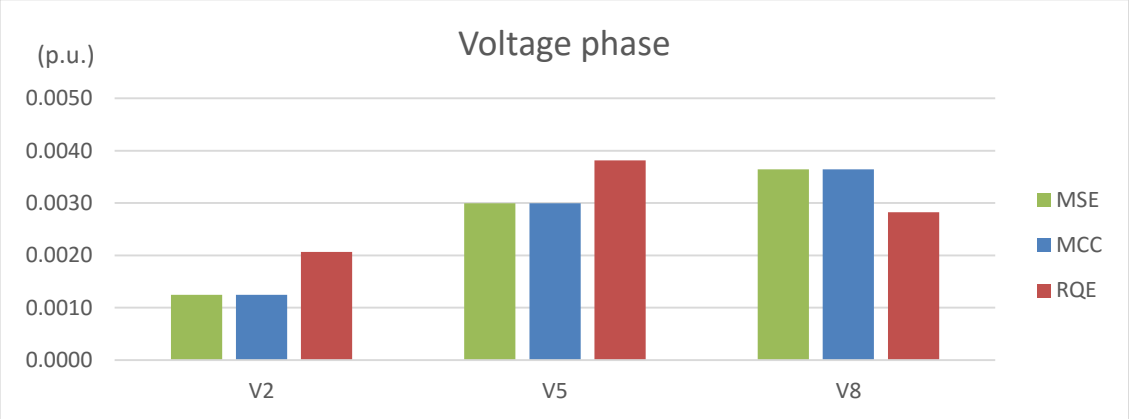

Figure 6-11- Absolute value of residual errors in voltage magnitudes for the AC model, without gross errors.

As it is possible to observe, the same conclusions of the DC model are transposed here. The MSE and MCC approach the same result, when there are no gross errors. The RQE leads to a different state with different residual errors. Despite of the difference, once the errors are always lower than the noise introduced to the original values, that state is still acceptable and reliable. Is still important to mention that, the point obtained with RQE is not worse than the one obtained by MCC or MSE. It's different, and in some components leads to a lower error, and in others to a higher error. In fact, it is the state obtained considering the criterion of less entropy, calculated by Renyi's Quadratic Entropy, and from that point of view it conducts to a better point, comparing to the ones obtained with the other criteria.

In the following chart is possible to see the errors distribution, calculated though the PW method with a standard deviation of 0.1.

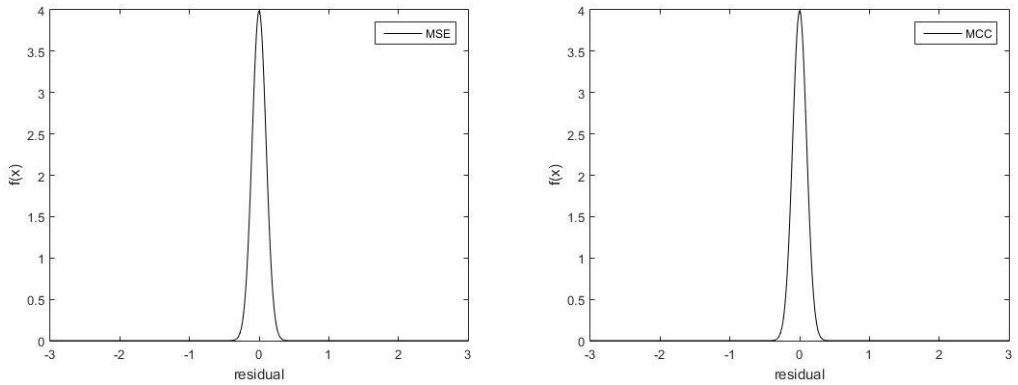

Figure 6-12 - pdf of the errors for the case without outliers, using MSE, on the left, and MCC on the right.

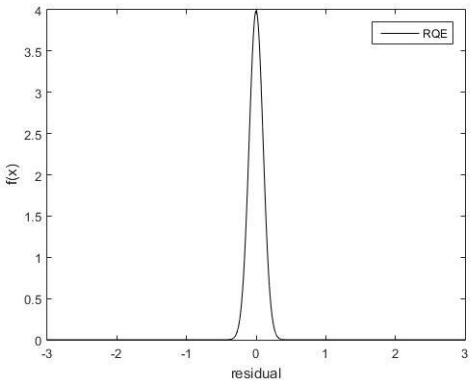

Figure 6-13- pdf of the errors for the case without outliers, using RQE.

In the previous tests, all the curves are similar of even equal, thus, the representation of each curve was made in independent charts. Therefore, a better observation of each curve is possible, without overlapping.

As there are no gross errors, and the noise introduced to the original measurements was a Gaussian cantered in zero, the residuals distribution assumed also that Gaussian form, for each one of the different metrics.

6.4 - State estimation using the fusion method and considering typical errors

In this section, two different cases of bad data are studied, in order to compare the information theory related concepts with the classic MSE. The cases are: one missing measurement and an inversion of line power flow, where both active and reactive power are inverted. For both cases the PW applied for MCC and RQE criteria had size 0.1.

6.4.1 - Missing measurement in P_{inj_6}

In this subsection is assumed an outlier, a typical error of a missing measurement. For that, the active injected power in bus 6 was assumed as zero. The state estimation was computed for the 3 different criteria, MSE, MCC and RQE.

The measurement set is the same of the table 6-2-measurements with gross errors (missing measurement).

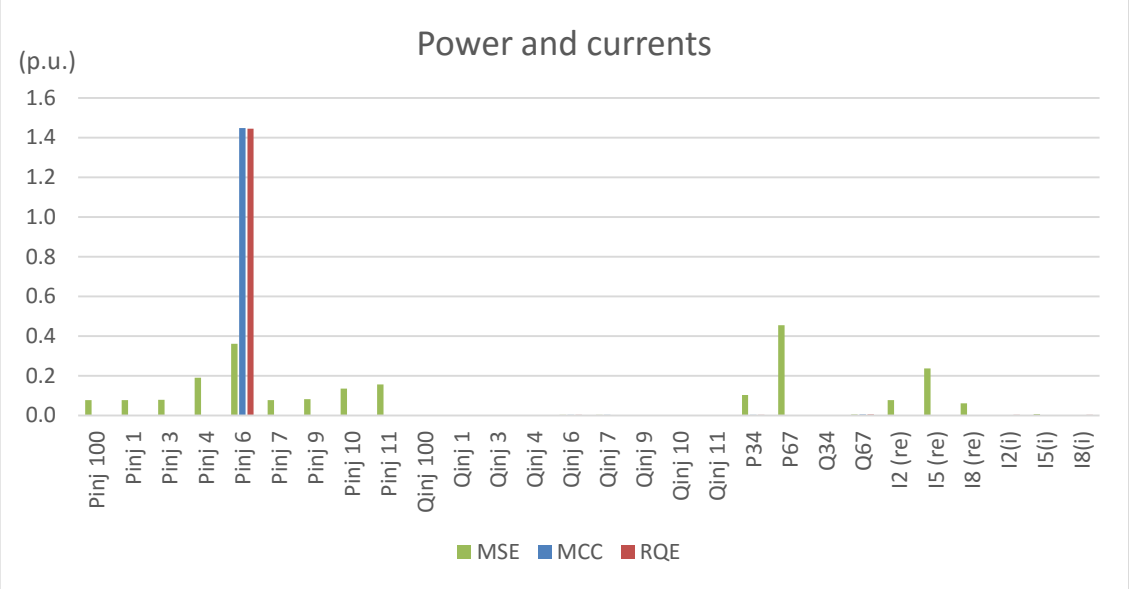

Figure 6-14 - Absolute value of residual errors in injected power and currents and lines power flow for the AC model with an missing measurement.

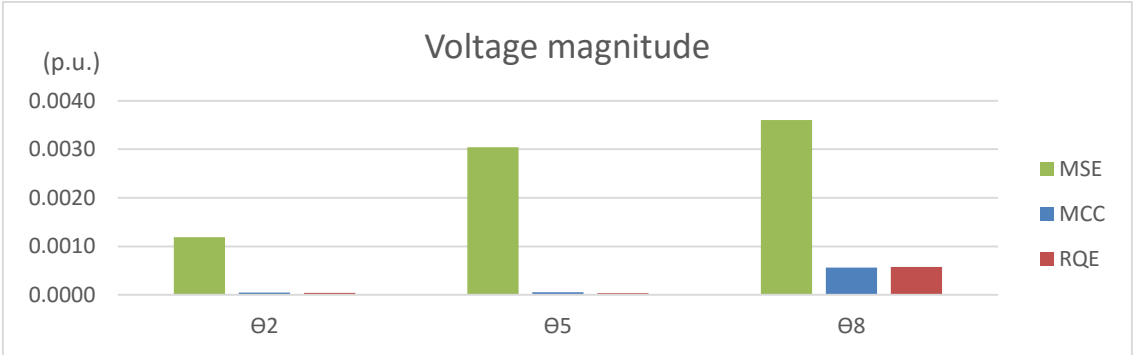

Figure 6-15 - Absolute value of residual errors in voltage phases for the AC model with an missing measurement.

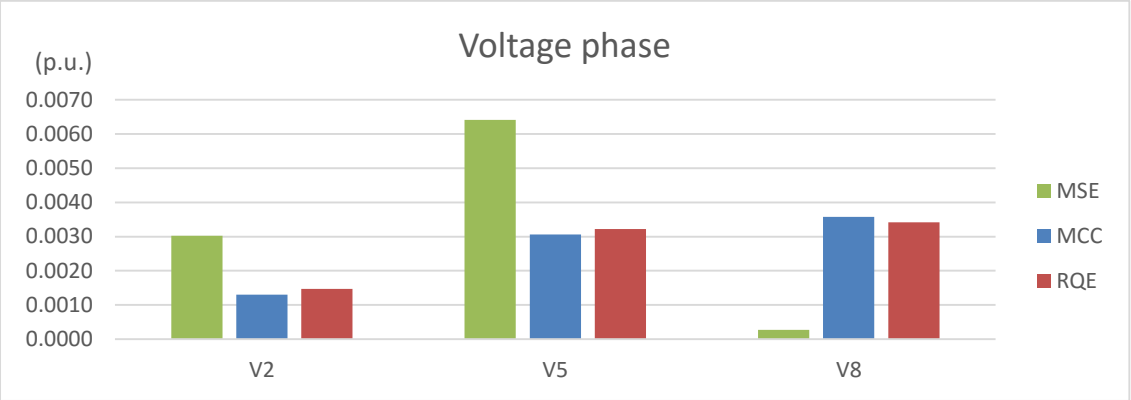

Figure 6-16- Absolute value of residual errors in voltage magnitudes for the AC model with an missing measurement.

From the analysis of the figures 6-15 to 6-17 there are two important aspects that require some attention. First, the criterions MCC and RQE provide really low errors to all the metrics, except one. The error on the $Pinj_6$ is extremely high. In fact, the estimated value by those criteria was approximately 1.45 p.u., and the measured value is zero. Despite the big residual error, it's known that the true error is really low, since the true component value is around 1.45 p.u., table 6-1. Thereby, these criteria were able to identify and ignore the component that contain an outlier. Second, the MSE generated bigger errors than the other metrics, for almost all the components. These results show that MSE is incapable to identify bad data, thus, the gross error present in one measurement infects the estimations on all the other components.

The distribution of the errors obtained using each metrics can be found in the following graphic charts. These graphics were obtained, as in the previous section, by the PW method, applying a Gaussian kernel with standard deviation 0.1 in each one of the residuals.

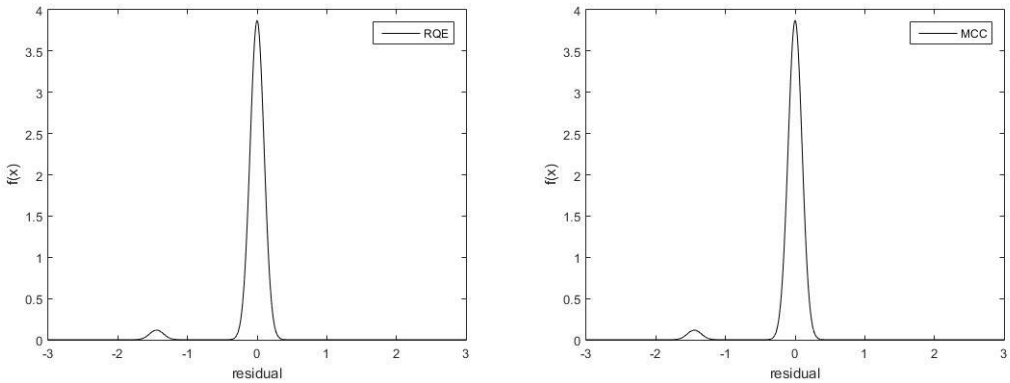

Figure 6-17 - pdf of the residual errors for the case of an missing measurement, using RQE, on the left, and MCC on the right.

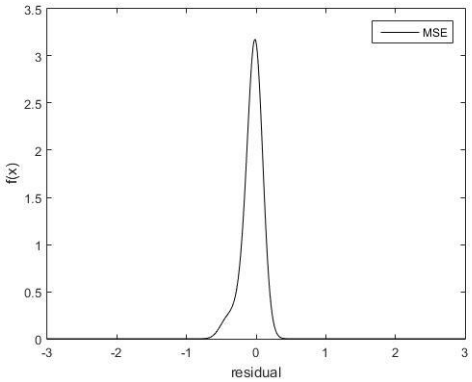

Figure 6-18 - pdf of the residual errors for the case of an missing measurement, using MSE.

From the analysis of the errors distributions, figures 6-18 and 6-19, it's easily understood the explained in the previous paragraphs. The metrics MCC and RQE, figure 6-18, can identify perfectly the outlier. For that reason, it has a high Gaussian centered in zero, and then, a peak in -1.45. This peak represents the error on the estimation of the component affected with the outlier. It means that only the error on that component estimation is high, which is the intended. The true measurement value should be 1.45 p.u. and not 0, thus, the error is -1.45 p.u., which is the value identify by the MCC and RQE.

The same does not happen with MSE, figure 6-19. That point is not identified, and is not ignored. The consideration of that measurement as a reliable one had a bad effect in other components' estimation. Therefore, the curve obtained with MSE were distorted and the left tail of the Gaussian was extended in the direction of the outlier.

6.4.2 - Inversion of the power flow.

In this subsection another typical error was studied. The inversion of a line power flow. For that the signal of the active and reactive power of the line 6-7 were changed.

The set of used measurements is presented in table 6-2- measurement affected with gross errors, inverse power flow.

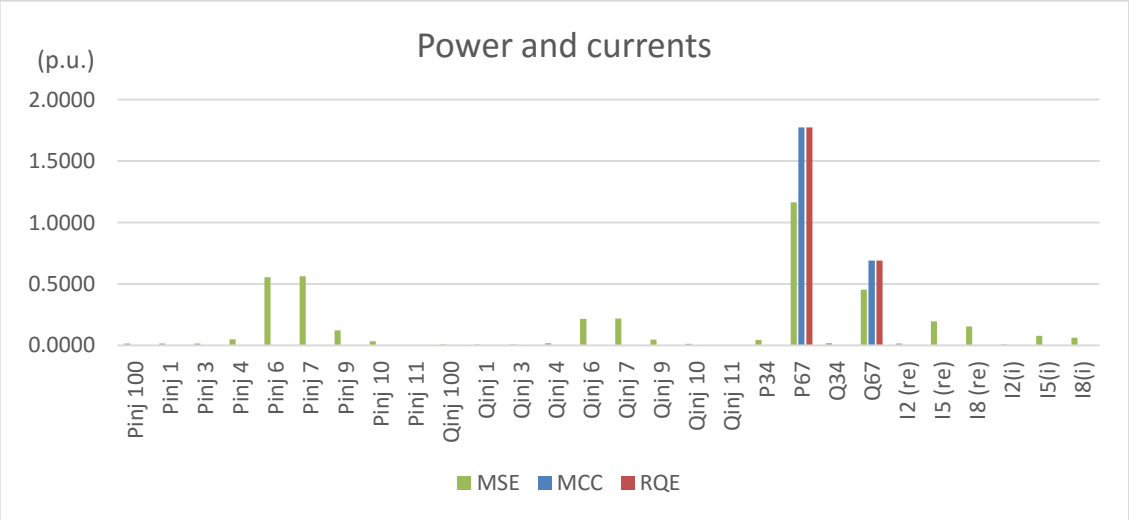

Figure 6-19 - Absolute value of residual errors in injected power and currents and lines power flow for the AC model with inversion on one line power flow.

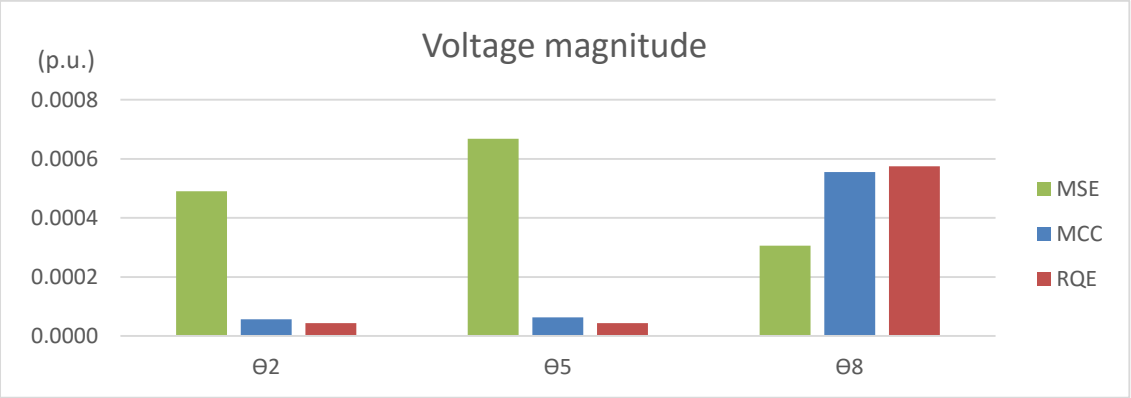

Figure 6-20 - Absolute value of residual errors in injected power and currents and lines power flow for the AC model with inversion on one line power flow.

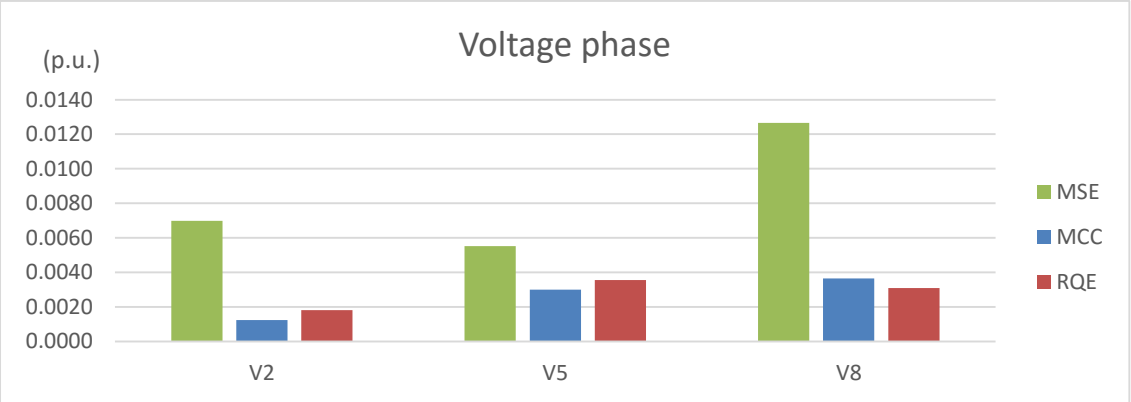

Figure 6-21 - Absolute value of residual errors in injected power and currents and lines power flow for the AC model with inversion on one line power flow.

In resemblance of what happens in the previous subsection, the errors obtained with MCC and RQE assume really high values for the components affected with a large error, $P_{6\text{-}7}$ and $Q_{6\text{-}7}$, and low values to all the other components. Another important aspect is that the residuals in $P_{6\text{-}7}$ and $Q_{6\text{-}7}$ is approximately the double of the measured value, what is the intended. It means that these criteria were able to identify the components affected with gross errors and ignore them. The estimated values, for these components, were symmetrical to their measurements, that is, the values that are known to be the true ones.

The measurements were obtained from a power flow result, introducing some noise to that values, more information about this process can be found in annex A.3. The minimization of MSE conducted to a state with big errors in all the components, generating estimation errors higher than the noise introduced to the original values (power flow results).

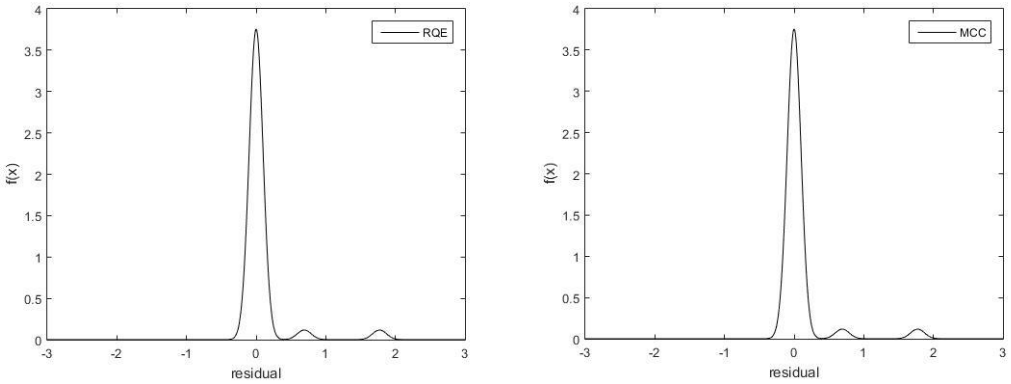

Figure 6-22 - pdf of the residual errors for the case of an inversion of a line power flow, using RQE, on the left, and MCC on the right.

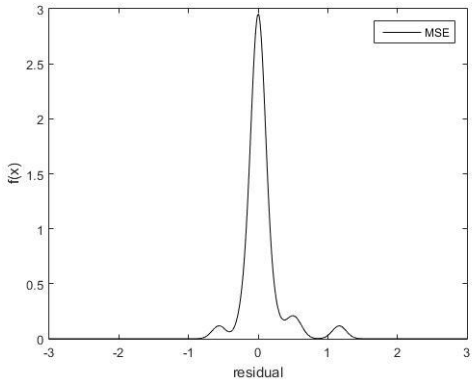

Figure 6-23 - pdf of the residual errors for the case of an inversion of a line power flow, using MSE.

In figure 6-23 is possible to observe a Gaussian centered in zero and two peaks, the peaks are localized in the values of the residual errors for the components affected with the measure error. This means that MCC and RQE were able to identify the outliers, conducting to a proper state estimation, where the gross errors didn't interfere. Even that, in this subsection, were considered two outliers, and not only one as in the previous subsection, the developed method was able to compute a reliable estimation.

The minimization of MSE, figure 6-24, didn't identify or ignored the outliers. For that reason, in figure 6-24 is not visible a Gaussian but a deformed curve, which represents the influence of the outliers in the error's distribution. Thus, it's easy to conclude that, using the conventional squared error, the outliers are not ignored and distort the estimations.

6.5 - Chapter conclusion

There are really important conclusions that have to be highlighted in this chapter. The first one is about the integration of PMU in the sensor system. In fact, the substitution of some conventional sensors by PMU improving the state estimation results, leading to lower errors than the ones obtained with only conventional sensors. Thus, it was established that the fusion of conventional sensors with PMU doesn't result in a catastrophic fusion, but in an improvement of the system's representation. This results can be explaining by the high precision of PMU, and also, by the inclusion of voltage phases to the measurement set. PMU has other interesting

advantages to the power system, among which: measurements synchronized by GPs-clock and lower sampling periods.

The developed fusion method was tested in a network based in real parameters, more specifically, based in a typical European medium voltage distribution network [29]. It is important to highlight the need for further studies about the stochastic fusion method, already proposed in the previous chapters. Such studies must provide information about each sensory system's probability to conduct the estimation to the real state. As that information is not yet available, the solution was to find the optimal point closer to the ideal one. Such ideal point is given by (max fit_1, max fit_2), as already discussed in chapter 5. In fact, being possible to simplify the objective function's expressions as they are in table 6-5, this ideal point doesn't need to be calculated when the fusion metric is L_1.

Lastly, but not less important, in this chapter was possible to demonstrate, in a network based on real parameters, that concepts related with information theory, namely MCC and RQE, have the ability to naturally identify and ignore outliers. These criteria provide estimations as if the measurements didn't contain outliers. That's a big achievement once these criteria can afford proper state estimations without data pretreatment. As it is well known, the MSE doesn't have that capability and, because of that, the state estimator provided with MSE, or WLS, requires additional tools to eliminate bad data.

Chapter 7

Conclusions and future work

This thesis had as main objective the development of a sensory fusion method capable to distinguish two sensory systems and to assign different trust values to each of them. Meanwhile, this method should integrate concepts related with information theory as an alternative to the traditional state estimation criterion, the weighted least square. During the development of this thesis, many studies were made to find its advantages when compared to the traditional method.

The main conclusions and future work, based on what was developed, are presented in the following sections.

7.1 - Conclusions

The development of this thesis allowed the development and application of a new sensory fusion method. Such method is able to integrate difference sensor classes in the same measurement system, and to consider their information separately. It was also intended to apply alternative concepts as state estimation criteria. These concepts should be able to naturally identify and discard measurements corrupted by gross errors.

Therefore, in a first study case, innovative concepts related with information theory were applied as the state estimation optimization function. Both metrics, Correntropy induced metric (also referred in this work as the Maximum Correntropy Criterion) and Renyi's quadratic entropy, were tested in a non-resistive network for simple studies and then applied to a network with benchmark parameters of a European medium voltage network, respectively in a DC model system and in an AC model. In both cases, these metrics, MCC and RQE, were compared with the traditional least square error (LSE) presenting a similar behavior when the measured data is free of outliers. It is important to note that, when the measured values contain gross errors, the MCC and the RQE criteria can identify and ignore them. This way, these metrics provide a proper state estimation. These results may be easily explained by MCC's behavior with the increasing distance between two points, transiting from L2, L1 to L0 norms. This behavior is well described in Chapter 2. For the RQE case, the explanation is based on the perspective of entropy minimization of the errors distribution, thereby, an error displaced from the errors distribution is automatically ignored. Both MCC and RQE behaviors depend on the

selected Parzen window size, thus, a wider window allows the assumed error distribution to be more embracing than a narrower window. A thinner window is more restrict, allowing the outliers to be identified and ignored.

In Chapter 6 is possible to observe the SE's results in the presence of two usual cases of outliers, namely missing measurement and inversion of a line power flow. While these outliers notably influenced the state estimation provided by the minimization of MSE, leading to high deviations for most of the components, MCC and RQE managed to provide good estimations for all of the components including those affected by the error.

The study about the inclusion of the PMU in the measurement system proved to be important as it leads to a better state estimation, by minimizing its residual errors. Such results can be explained by the high trust in PMU, and also, by the inclusion of voltage phases to the measurement set. PMU has other interesting advantages to the power system, among which: measurements synchronized by GPs-clock and lower sampling periods.

The big progress provided by this thesis is the perspective assigned to the fusion method of two separately sensory systems. A perspective of stochastic fusion was born here. However, further research and tests must be done in order to find the stochastic values to assign to each sensory system.

For now, such information is not available yet, hence, this work used another way to choose the optimal fusion point. This point was selected by its minimal distance to an ideal point. The ideal point is the point with coordinates defined by both best fitness evaluations of each sensory system. The chosen metric to evaluate that distance was the absolute norm. The reason for this choice was its simplicity and its capability to be easily converted into a stochastic fusion. Such transition will be possible when there's a larger amount of information on the probability of each sensory system to provide the estimation corresponding to the real state.

The fusion method doesn't interfere with the properties of the evaluation functions MCC and RQE. In fact, in Chapter 6 it is possible to behold the application of the fusion method using those evaluation metrics in a real parameters based network. The results are convenient and show good properties when facing outliers.

Summing up: in this thesis an innovative fusion method, integrating also alternative state estimation criteria, was developed and proposed. The developed method was tested in different network circumstances presenting always advantages in comparison with the traditional process.

7.2 - Future work

The evident objective of this thesis was to expose promising properties of the developed method and to confirm its advantages related to the conventional process. For that proof of concept, a meta-heuristic algorithm – EPSO - was used.

As such properties were presented and its convenient application was proven, now is desirable to adapt the developed method to real time problems. The EPSO algorithm takes a lot of time to find the optimal fusion point in real networks. Thus, the need for a development of a real-time resolution method emerges. One proposal for a future work is to resolve the proposed optimization problems by a differential method such as the Gauss-Newton.

Another imperative work is about the stochastic fusion. As it was shown in Chapter 4, the state estimations' results differ a lot with the variation of the weights assign to each one of

the sensory system. To choose the state closer to the real one, a stochastic fusion may be used, where the parameters α and ß, described in Chapter 4, are determined by the probability of each sensory system to lead to the real system's state. To obtain such information, studies must be done considering: reliability of each sensory type, the number of sensors in each system, their location and other possible issues as errors by asynchronism.

One determining factor for the outlier's identification is the Parzen window size. In the previous studies it was possible to find a parameter able to provide a proper identification and treatment of bad data. However, the selection of that parameter has to take in consideration the magnitude of the errors. More studies are required to elaborate a process able to determine the ideal value for the Parzen window's size according to the environment where it will be inserted.

References

1. V. Miranda, A. Santos, and J. Pereira, “State Estimation Based on Correntropy: A Proof of Concept,” *in IEEE Trans. Power Syst.*, vol.24, no.4, pp.1888,1889, Nov. 2009.
2. Costa, Antonio Simões, André Albuquerque, and Daniel Bez. "An estimation fusion method for including phasor measurements into power system real-time modeling", *in IEEE Transactions on Power Systems*, vol.28, no.2 , pp.1910-1920, 2003.
3. Zeleny, M. " Multiple Criteria Decision Making, McGraw-Hill", New York,1982.
4. C.E. Shannon, "A mathematical theory of communication", *Bell System Technical Journal*, Vol. 27, no.3, pp. 379–423, 1948.
5. J. C. Principe, " Information theoretic learning: Renyi's entropy and kernel perspectives", Springer Information and Science, Springer-Verlag New York, 2010, Chapters 1 and 2.
6. E. Parzen, “On the estimation of a probability density function and the mode”, Annals Math. Statistics, vol. 33, 1962, p. 1065.
7. V. Miranda, "Information Theoretic Learning Principles a short tutorial", in ISAP Conference and debate, 2015, pp.12-15.
8. W.F.Liu, P.P. Pokharel, and J.C. Principe, "Correntropy: A localized similarity measure", in 2006 IEEE International Joint Conference on Neural Network Proceedings, 2006: pp. 4919-4924.
9. Liu, W.F., P.P. Pokharel, and J.C. Principe," Correntropy: properties and applications in non-gaussian signal processing", in IEEE Transactions on Signal Processing, 2007. vol.55, no.11, pp. 5286-5298.
10. Clara Gouveia, A.S., Helder Leite, "The Application Of Distribution State Estimation To Support A Real-Time Voltage Control Algorithm: A Path To Increase The Integration Of Distributed Generation", *in* 21st International Conference on Electricity Distribution, 2011.
11. André Madureira, R.B., Luís Seca, "Advanced System Architecture And Algorithms For Smart Distribution Grids: The Sustainable Approach", *in* 23rd International Conference on Electricity Distribution, 2015.
12. Paula S. Castro Vide, FP Maciel Barbosa, and Isabel M. Ferreira, "State estimation model including synchronized phasor measurements." in *Universities' Power Engineering Conference (UPEC), Proceedings of 2011 46th International*. VDE, 2011.
13. Paula S. Castro Vide, FP Maciel Barbosa, and Isabel M. Ferreira, "Combined use of SCADA and PMU measurements for power system state estimator performance

enhancement." in *Energetics (IYCE), Proceedings of the 2011 3rd International Youth Conference on*. IEEE, 2011.

14. A. Monticelli, "State Estimation In Electric Power Systems, A Generalized Approach", Kluwer Academic Publishers , Chapter 2, 1999.
15. W. Wu, Y.G., B. Zhang, A. Bose, S. Hongbin," Robust state estimation method based on maximum exponential square", *in* IET Generation, Transmission & Distribution, 2011.
16. Abur, Ali, and Antonio Gomez Exposito, "Power system state estimation: theory and implementation", CRC press, 2004, Chapter 2.
17. *M-estimators* by Zhengyou Zhang , 1996. Available on http://research.microsoft.com/en-us/um/people/zhang/INRIA/Publis/Tutorial-Estim/node24.html , last acess: 19th June 2016
18. Green, Peter J. "Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives", in *Journal of the Royal Statistical Society. Series B (Methodological)* , pp:149-192,1984.
19. W.J.J. Rey, "Introduction to Robust and Quasi-Robust Statistical Methods", ed. Springer. Berlin, Heidelberg 1983, Chapter 8.
20. Mitchell, H.B., "Multi-sensor data fusion an introduction", New York: Springer, 2007, Chapter 1.
21. K. Faceli, A.C.P.L.F. De Carvalho, and S.O. Rezende, "Combining intelligent techniques for sensor fusion", Applied Intelligence, vol.20, no.3: pp. 199-213, 2004.
22. G., K., "Data Fusion: From primary metrology to process measurement", *in IEEE Instrumentation and measurement*, vol.3 pp:1325-1329, 1999.
23. Karthik Nandakumar, "Integration of multiple cues in biometric systems", Diss. Michigan State University, pp:1-11, 2005.
24. R. Boudjemaa, F.A.B.," *Parameter estimation methods for data fusion*", National physical Laboratory Report No. CMSC 38-04, 2004.
25. F., D.-W.H., "Sensor models and multisensor integration", in *International Journal of Robotics Research no.7, pp.97-113*. 1988.
26. Belur V. Dasarathy, "Decision fusion". Vol. 1994. Los Alamitos, CA: IEEE Computer Society Press, 1994.
27. C. A. Noonan, and M. Pywell. "Aircraft sensor data fusion: An improved process and the impact of ESM enhancements", *in Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar* (1998).
28. P. N. Pereira Barbeiro, et al. "Exploiting autoencoders for three-phase state estimation in unbalanced distributions grids" *Electric Power Systems Research* 123 pp:108-118, 2005
29. Kai Strunz, N. Hatziargyriou, and C. Andrieu. "Benchmark systems for network integration of renewable and distributed energy resources", *Cigre Task Force C*6 (2014): 04-02, pp:33-37
30. V. Miranda, H. Keko, and A. J. Duque, "Stochastic star communication topology in evolutionary particle swarms (EPSO)", Int. J. Comput. Intell. Res., vol. 4, no. 2, pp.105-116, 2008.
31. Opara, Jakov Krstulovic. "Information Theoretic State Estimation in Power Systems." (2014).

32. Aminifar, Farrokh, et al. "Synchrophasor measurement technology in power systems: Panorama and state-of-the-art", *in IEEE Access* 2 (2014): 1607-1628.

Annex A

12 busses system

The network used to test the different fusion methods, and state estimation criteria, is presented in this chapter. This network consists in an adaptation from a Medium Voltage Distribution Network Benchmark, European configuration, from Cigre reports [29]. A typical European MV distribution network is three-phase and its structure can either be meshed or radial, being that in rural environment it tends to be radial.

There is a big effort to keep the European networks balanced. Once those networks are assumed to be symmetric and balanced, in the study cases, the neutral wires are going to be ignored. Therefore, the presented line parameters will be already the equivalent of the three phases.

As it is possible to see in figure A-1 there are some distributed energy resources (DER) along the network. More specifically, in buses 1, 6 and 9, represented by generators in the scheme of figure A-1. In the next sessions is possible to see the network topology, line parameters, the DER productions, load characteristics, injected power, the true values power flow values and "measurements" affected by Gaussian errors with small dimension.

A.1- Topology

The topology of the utilized network is shown in figure A-1. In the connection from bus 100 to 1, is possible to see a transformer with voltage 63/15 KV with a leakage reactance of 8% and a nominal power 25MVA.

The branches impedances can be consulted in the following table, A-1, measured in Ω, and in table A-2, measured in p.u. To convert into p.u. system it was used the base power S_b=25 MVA to the DC model and S_b=2.5 MVA to the AC model. The voltage base is Vb=15KV for the medium voltage part of the network (buses 1 to 11) and V_b=63KV for the bus 100.

94 12 buses sytem

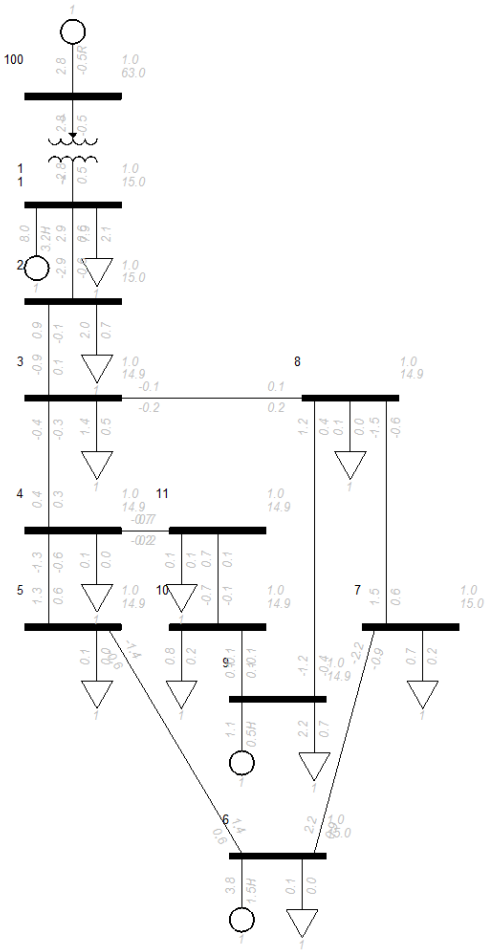

Figure A-1-Network topology.

Table A-1 Line parameters in ohm.

Branch	From	To	R	X
			(Ω)	(Ω)
1	1	2	0.31640	0.35000
2	2	3	0.49720	0.55000
3	3	4	0.06780	0.07500
4	4	5	0.06780	0.07500
5	5	6	0.16950	0.18750
6	6	7	0.02260	0.02500
7	7	8	0.19210	0.21250
8	8	9	0.03390	0.03750
9	9	10	0.09040	0.10000
10	10	11	0.03390	0.03750
11	11	4	0.05650	0.06250
12	3	8	0.14690	0.16250

Table A-2- Line parameters in p.u.

Branch	From	To	R (p.u.) with Sb=25MVA	X (p.u.) with Sb=25MVA	R (p.u.) with Sb=2.5MVA	X (p.u.) with Sb=2.5MVA
1	1	2	0.035156	0.038889	0.003516	0.003889
2	2	3	0.055244	0.061111	0.005524	0.006111
3	3	4	0.007533	0.008333	0.000753	0.000833
4	4	5	0.007533	0.008333	0.000753	0.000833
5	5	6	0.018833	0.020833	0.001883	0.002083
6	6	7	0.002511	0.002778	0.000251	0.000278
7	7	8	0.021344	0.023611	0.002134	0.002361
8	8	9	0.003767	0.004167	0.000377	0.000417
9	9	10	0.010044	0.011111	0.001004	0.001111
10	10	11	0.003767	0.004167	0.000377	0.000417
11	11	4	0.006278	0.006944	0.000628	0.000694
12	3	8	0.016322	0.018056	0.001632	0.001806

A.2- Load and DER data

The network loads are dived into two groups, “residential” and “commercial/industrial”. In the next table, A-3, is possible to see the maximum value for the apparent power of the loads in each bus, as well as the power factor of each installation type.

Table A-3 - Apparent power and power factor for the loads in each bus.

Bus	Apparent Power, S (kVA)		Power factor	
	Residential	**Commercial/Industrial**	**Residential**	**Commercial/Industrial**
1	8700	4500	0,98	0,95
2		2500		0,95
3	185	1650	0,98	0,95
4	245		0,98	
5	250		0,98	
6	265		0,98	
7		900		0,95
8	305		0,98	
9		2750		0,95
10	290	800	0,98	0,95
11	170		0,98	

In figure A-1 is represented some DER in buses 1, 6 and 9. Such distributed resources are cogeneration, mini-hydric and solar photovoltaic as represented in table A-4.

Table A-4 - Maximum power generated by each DER.

Bus	Technology	Active Power, P (kW)	Reactive power, Q (kvar)	
			Min	Max
1	cogeneration	8000,0	0	3200,0
6	Mini-hydric	5000,0	0	2000,0
9	Solar photovoltaic	3000,0	0	1200,0

As it is obvious, the loads are not at the maximum value all time, or at the same time. Thus, in table A-5 is exposed the loads profile, as well as DER productions profiles, according to the time of the day. To provide the experiments for this work, it was chosen the hour 17, as it is highlighted in table A-5.

Table A-5- Power distribution according time.

Time (h)	Residential load (%)	Industrial load (%)	cogeneration (%)	Mini-hydric (%)	Solar (%)
0	30,0%	10,0%	0,0%	30,0%	0,0%
1	27,5%	12,0%	0,0%	30,0%	0,0%
2	25,0%	12,0%	0,0%	30,0%	0,0%
3	22,5%	15,0%	0,0%	30,0%	0,0%
4	20,0%	15,0%	0,0%	30,0%	0,0%
5	20,0%	15,0%	0,0%	30,0%	0,0%
6	22,5%	18,0%	0,0%	30,0%	2,9%
7	30,0%	25,0%	0,0%	30,0%	14,9%
8	40,0%	50,0%	50,0%	75,0%	35,2%
9	43,0%	80,0%	50,0%	75,0%	54,8%
10	46,0%	100,0%	100,0%	75,0%	70,2%
11	50,0%	100,0%	100,0%	75,0%	81,4%
12	50,0%	90,0%	100,0%	75,0%	87,5%
13	55,0%	50,0%	75,0%	75,0%	88,6%
14	60,0%	50,0%	75,0%	75,0%	84,5%
15	60,0%	94,0%	100,0%	75,0%	75,6%
16	55,0%	90,0%	100,0%	75,0%	60,1%
17	50,0%	85,0%	100,0%	75,0%	38,2%
18	65,0%	70,0%	75,0%	75,0%	16,6%
19	85,0%	40,0%	50,0%	75,0%	3,5%
20	100,0%	30,0%	0,0%	75,0%	0,1%
21	90,0%	10,0%	0,0%	30,0%	0,0%
22	75,0%	10,0%	0,0%	30,0%	0,0%
23	55,0%	10,0%	0,0%	30,0%	0,0%

A.3- Power flow results and measurements

The values presented previously are generic values for the network. However, it's necessary only one case study to provide some experiments in this work. For that, the time 17 a.m. was selected, being the loads on each bus presented, for that hour, in table A-6, in MVA. In table A-7 are represented the generations by the multiple DER, either in MW or Mvar.

Table A-6- Loads in hour 17 a.m.

Bus	Load	
	P (MW)	Q (Mvar)
1	7.89675	2.059995
2	2.01875	0.663531
3	1.423025	0.456338
4	0.12005	0.024377
5	0.1225	0.024875
6	0.12985	0.026367
7	0.72675	0.238871
8	0.14945	0.030347
9	2.220625	0.729884
10	0.7881	0.241185
11	0.0833	0.016915

Table A-7- Generated Power by the DER in the at 17a.m.

Bus	Generated Power	
	P (MW)	Q (Mvar)
1	8	3.2
6	3.75	1.5
9	1.146	0.4584

Table A-8- Injected Power in MW and Mvar at 17 a.m.

Bus	Injected Power	
	P (MW)	Q (Mvar)
100	2.802903	-0.59758
1	0.10325	1.140005
2	-2.01875	-0.66353
3	-1.42303	-0.45634
4	-0.12005	-0.02438
5	-0.1225	-0.02487
6	3.62015	1.473633
7	-0.72675	-0.23887
8	-0.14945	-0.03035
9	-1.07463	-0.27148
10	-0.7881	-0.24118
11	-0.0833	-0.01691

Table A-9- Injected Power in p.u. at 17 a.m.

Bus	Injected Power		Injected Power	
	P (p.u.) with Sb=25MVA	Q (p.u.) with Sb=25MVA	P (p.u.) with Sb=2.5MVA	Q (p.u.) with Sb=2.5MVA
100	0.1121	-0.0239	1.1212	-0.2390
1	0.0041	0.0456	0.0413	0.4560
2	-0.0808	-0.02654	-0.8075	-0.2654
3	-0.0569	-0.01825	-0.5692	-0.1825
4	-0.0048	-0.00098	-0.0480	-0.0098
5	-0.0049	-0.00099	-0.0490	-0.0099
6	0.1448	0.058945	1.4481	0.5895
7	-0.0291	-0.00955	-0.2907	-0.0955
8	-0.006	-0.00121	-0.0598	-0.0121
9	-0.043	-0.01086	-0.4299	-0.1086
10	-0.0315	-0.00965	-0.3152	-0.0965
11	-0.0033	-0.00068	-0.0333	-0.0068

What really matter for these studies is not the consume or generation, but the injected power at each bus. The injected power at each bus, in p.u., can be consulted in table A.8.

A.3.1 AC model measurements

With the assistance of the program PSS/E was possible to solve the power flow. It was obtained the voltage magnitude and phase on each bus, according to the load and distributed generation. These results can be consulted in table A-10.

Table A-10- Voltage profile at 17 a.m.

Bus Number	Voltage (p.u.)	Angle (degrees)	Angle (rad)
1	1.001952	-0.512908	-0.0089519
2	0.997080	-0.730551	-0.0127505
3	0.995537	-0.874560	-0.0152640
4	0.995799	-0.873273	-0.0152415
5	0.996383	-0.858845	-0.0149897
6	0.997958	-0.818048	-0.0142776
7	0.997639	-0.827246	-0.0144382
8	0.995781	-0.877771	-0.0153200
9	0.995537	-0.886272	-0.0154684
10	0.995442	-0.887641	-0.0154923
11	0.995566	-0.882662	-0.0154054
100	1.000000	0.000000	0.0000000

It's important to mention that, for the application of the AC model, the base power selected was 2.5MVA. The reason why it was selected a different value from the one selected in the DC model, and also different from the transformer nominal power, was to make more concluding tests in the sixth chapter.

The values selected as measurements (without PMU) are presented in the table A-11. These measurements were generated from original values resulted from the power flow, shown in the same table, affected by:

- Gaussian error with the width [-0.005; 0.005] in the injected power;
- Gaussian error with the width [-0.004; 0.004] in the voltage magnitude obtained from PSS/e, table A-10.

Table A-11-Power flow results and generated measurements without PMU.

	Bus	Power Flow result	Measurement affected with noise
P_{inj}	100	1.121140	1.122810
	1	0.041322	0.039534
	2	-0.807506	-0.809074
	3	-0.569125	-0.572176
	4	-0.048778	-0.045104
	5	-0.048648	-0.043448
	6	1.449080	1.451090
	7	-0.291667	-0.291622
	8	-0.059664	-0.062184
	9	-0.430091	-0.429711
	10	-0.315175	-0.318894
	11	-0.032993	-0.033324
Q_{inj}	100	-0.238982	-0.236943
	1	0.455981	0.457423
	2	-0.265455	-0.267485
	3	-0.182535	-0.182715
	4	-0.010337	-0.011612
	5	-0.009668	-0.009689
	6	0.590705	0.588743
	7	-0.096752	-0.094147
	8	-0.011773	-0.010499
	9	-0.109040	-0.109329
	10	-0.096427	-0.098370
	11	-0.006464	-0.006126
V	2	0.997079	0.998354
	5	0.996383	0.999383
	8	0.995781	0.992141

The injected power shown in table A-11 was obtained by applying the results of PSS/E (table A-10) to the power injections equations

$$P_i = V_i \sum_{j \epsilon Nj} V_j (G_{ij} cos\theta_{ij} + B_{ij} sin\theta_{ij}) \text{ and} \quad \text{(A.1)}$$

$$Q_i = V_i \sum_{j \epsilon Nj} V_j (G_{ij} sin\theta_{ij} - B_{ij} cos\theta_{ij}) . \quad \text{(A.2)}$$

In the previous measurement set are considered that each bus has power injection sensors, and that in buses 2, 5 and 8 are voltage magnitude sensors.

It was assumed, for this work, that also voltage magnitude and phases would be measured because of the introduction of PMU to the network, table A-12. The localization of the PMU was selected randomly. The chosen buses were 2, 5 and 8. therefore, these buses have, not only, injected current (phase and magnitude) measurements, but also the voltage phase and magnitude measurements. Thus, injected power and voltage magnitude sensors on buses 2, 5 and 8 were replaced by PMU. The measurements obtained for that buses injected currents and voltage magnitude and phases can be found in table A-13.

The injected current values where obtained by the following expression

$$Ii = \sum_{j \epsilon Nj} y_{ij} V_j , \quad \text{(A.3)}$$

the active current by

$$Ii(real) = \sum_{j \epsilon Nj} V_j (G_{ij} \cos(\theta_j) + B_{ij} sin(-\theta_j)), \quad \text{(A.4)}$$

and the reactive current by

$$Ii\ (imaginary) = \sum_{j \epsilon Nj} V_j (G_{ij} \sin(-\theta_j) - B_{ij} \cos(-\theta_j)). \quad \text{(A.5)}$$

Table A-12 -Power flow results and generated measurements with PMU, for injected power.

	Bus	Power Flow result	Measurement affected with noise
P_{inj}	100	1.12114	1.12281
	1	0.041322	0.039534
	3	-0.569125	-0.572176
	4	-0.048778	-0.045104
	6	1.44908	1.45109
	7	-0.291667	-0.291622
	9	-0.430091	-0.429711
	10	-0.315175	-0.318894
	11	-0.032993	-0.033324
Q_{inj}	100	-0.238982	-0.236943
	1	0.455981	0.457423
	3	-0.182535	-0.182715
	4	-0.010337	-0.011612
	6	0.590705	0.588743
	7	-0.096752	-0.094147
	9	-0.10904	-0.109329
	10	-0.096427	-0.09837
	11	-0.006464	-0.006126

The noise introduced to the injected power is the same already applied in table A-11. The noises introduced in voltage phases and current injections were, respectively, Gaussian with the width [-0.00005; 0.00005] and [-0.00004; 0.00004].

Table A-13 - Power flow results and generated measurements with PMU, for injected current and bus voltages.

	Bus	Power Flow result	Measurement affected with noise
teta	2	-0.0127503	-0.012826
	5	-0.0149895	-0.015083
	8	-0.0153199	-0.014795
V	2	0.997079	0.998354
	5	0.996383	0.999383
	8	0.995781	0.992141
I_{inj} (real) (p.u.)	2	-0.806411	-0.806061
	5	-0.0486732	-0.048913
	8	-0.0597286	-0.059329
I_{inj} (im) (p.u.)	2	-0.276537	-0.276589
	5	-0.0104342	-0.010506
	8	-0.0127391	-0.012667

To guarantee a better observability of the system, was added some lines power flows to the previous measurements. The lines power flows were calculated according to the following expressions

$$P_{ij} = V_{ij}^2(g_{si} + g_{ij}) - V_i V_j(g_{ij}cos\theta_{ij} + b_{ij}sin\theta_{ij}) , \quad \text{(A.6)}$$

$$Q_{ij} = -V_{ij}^2(b_{si} + b_{ij}) - V_i V_j(g_{ij}sin\theta_{ij} - b_{ij}cos\theta_{ij}), \quad \text{(A.7)}$$

and can be consulted in table A-14.

Table A-14- Measurements of lines power flows to the AC model

	line	Power Flow result	Measurement affected with noise
P_{ij} (p.u.)	3-4	0.0170434	0.0167534
	6-7	-0.0886902	-0.0886002
Q_{ij} (p.u.)	3-4	0.015892	0.015972
	6-7	-0.0344342	-0.0346142

A.3.2 DC model measurements

Once the previous network has a strong resistive component, it was made a simplification in that network in order to have measurements compatible to the DC model. thus, for the same network another study was made in PSS/E, but this time considering all the lines resistances as zero. The new results can be found in the table A-15.

Table A-15- PSS/E results considering R=0

Bus Number	Voltage (p.u.)	Angle (deg)	Angle (rad)
1	1.001954	-0.509292	-0.0088888
2	1.001163	-0.765749	-0.0133648
3	1.001558	-0.886939	-0.0154800
4	1.001690	-0.878788	-0.0153377
5	1.001886	-0.854360	-0.0149114
6	1.002397	-0.787507	-0.0137446
7	1.002302	-0.801535	-0.0139894
8	1.001719	-0.881691	-0.0153884
9	1.001659	-0.893212	-0.0155895
10	1.001619	-0.896662	-0.0156497
11	1.001644	-0.890455	-0.0155414
100	1.000000	0.000000	0.0000000

For the PSS/E results, once again, it was calculated the buses injected power, table A-15, That values were contaminated with noise just like in the previous subsection. The contamination was:

- Gaussian error with the width [-0.0004; 0.0004] to the power injections and lines power flows;
- Gaussian error with the width [-0.0005; 0.0005] to the voltage magnitude;
- Gaussian error with the width [-0.00002; 0.00002] to the voltage phase, considering the voltage phase error in radians.

The generated measurements can be found in table A-16 and A-17.

Table A-16-Power flow results and generated measurements, for active power, considering R=0.

	Bus	Power Flow result	Measurement affected with noise
P_{inj} (p.u.)	100	0.111326	0.111469
	1	0.00413	0.004325
	2	-0.080749	-0.080625
	3	-0.056928	-0.056763
	4	-0.004777	-0.004955
	5	-0.004907	-0.005229
	6	0.144789	0.144945
	7	-0.029052	-0.028692
	8	-0.005973	-0.006022
	9	-0.042991	-0.042779
	10	-0.031513	-0.031764
	11	-0.003354	-0.003434

Table A-17 - Power flow results and generated measurements, for reactive power and voltages, considering R=0.

	Bus	Power Flow result	Measurement affected with noise
Q_{inj} (p.u.)	100	-0.02393	-0.024226
	1	0.045606	0.045582
	2	-0.02654	-0.02667
	3	-0.018284	-0.018049
	4	-0.001043	-0.00102
	5	-0.000966	-0.000884
	6	0.058912	0.059035
	7	-0.009478	-0.009279
	8	-0.00133	-0.001663
	9	-0.010813	-0.010482
	10	-0.009614	-0.009353
	11	-0.000621	-0.000693
ϴ (rad)	2	-0.013365	-0.013383
	5	-0.015338	-0.014913
	8	-0.015388	-0.015403
V (p.u.)	2	1.001163	1.00148
	5	1.001886	1.00200
	8	1.001719	1.00214

To solve the state estimation by the DC model, it's necessary only few of the data presented in the previous table. Once the voltage magnitudes are assumed to be equal to 1 in all the buses, that components can be ignored. The same happens with the reactive injected power, once the voltage magnitudes are assumed to be close to one and the phase of contiguous buses are assumed to be nearly equal. Thereby, in table A-18 are presented the measurements used in the DC model.

Table A-18- DC measurements considering the adapted network. R=0.

	Bus/line	Power Flow result	Measurement affected with noise
P_{inj} (p.u.)	100	0.111326	0.111469
	1	0.004130	0.004325
	2	-0.080749	-0.080625
	3	-0.056928	-0.056763
	4	-0.004777	-0.004955
	5	-0.004907	-0.005229
	6	0.144789	0.144945
	7	-0.029052	-0.028692
	8	-0.005973	-0.006022
	9	-0.042991	-0.042779
	10	-0.031513	-0.031764
	11	-0.003354	-0.003434
ϴ (rad)	2	-0.013365	-0.013383
	5	-0.014911	-0.014913
	8	-0.015388	-0.015403

To guarantee observability of the network it was also introduced power flow measurements in lines 3-4 and 6-7. For the DC model, those components are easily calculated by the following expression

$$P_{ij} = \frac{\theta_i - \theta_j}{x_{ij}}. \quad \text{(X.8)}$$

where x_{ij} is the line ij reactance, and θ_i is the voltage phase at bus i.

Thus, to the measurements included in table A-16 can be added the power flow measurements, presented in table A-19. The calculated values were affected with a Gaussian error equivalent to the one induced in the injected power.

Table A-19- Power Flow measurements in lines 3-4 and 6-7 assuming DC model.

	Line	**Power Flow result**	**Measurement affected with noise**
P_{ij} (p.u.)	3-4	-0.017071	-0.016290
	6-7	0.088141	0.087880

It's important to mention that the voltage phases were obtained by PMU. In fact, the DC model will be used considering that there are PMU sensors on buses 2, 5 and 8. The PMU measures voltages and current phasors. For the simplifications of DC model is considered that the voltage magnitude is 1 p.u., thus, the injected power has the same value as the injected current. Hence, for the DC model will be consider that the PMU measurements are voltages phases and power injections (once for this model the power injections values are the same as current injections).

Annex B

Some results of the optimal fusion point

In this annex are present some graphics about the Chapter 5, section 3. This graphics allow a better observation of some results and also show some experimentations done until find the desirable result.

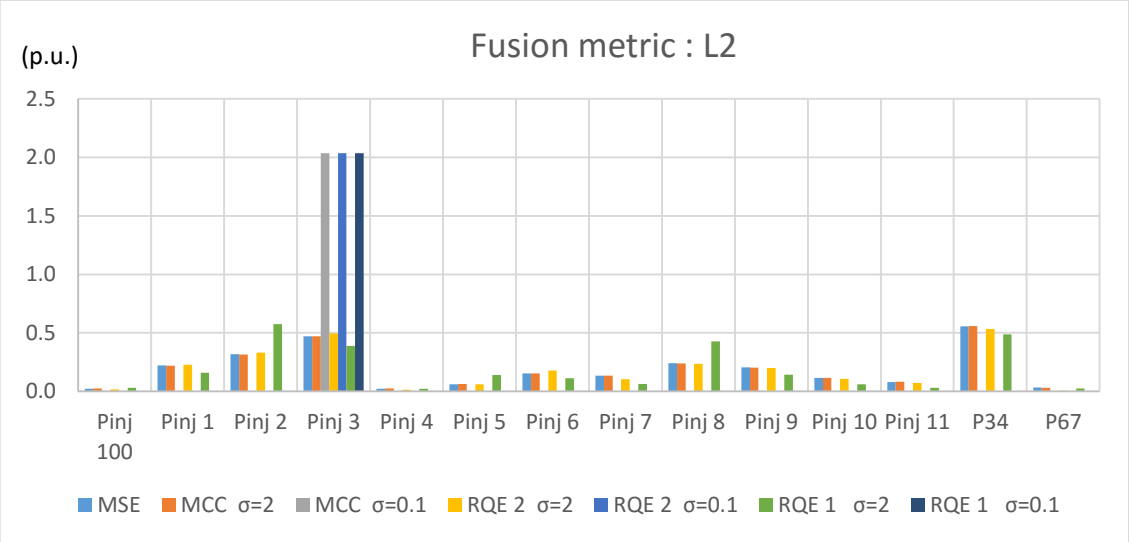

Figure B-1- Residual errors in buses power injections and lines power flows, using L2 as the fusion metric, with σ=2 and σ=0.1.

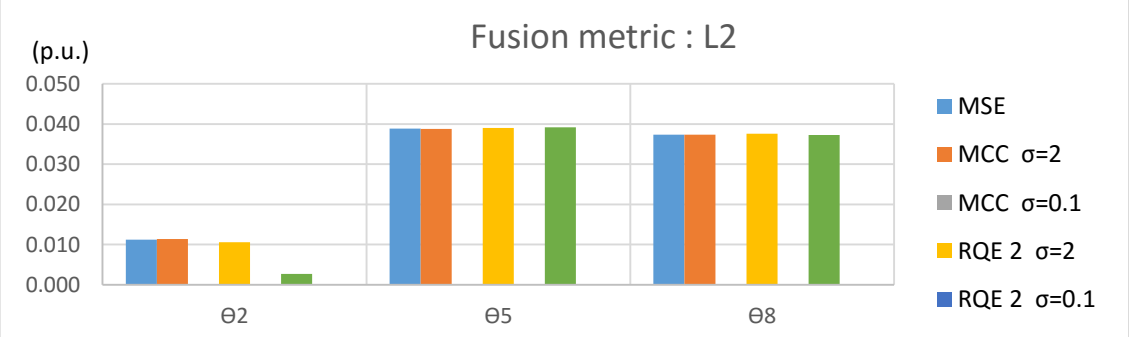

Figure B-2 - Residual errors in voltage phases using L2 as the fusion metric, with σ=2 and σ=0.1.

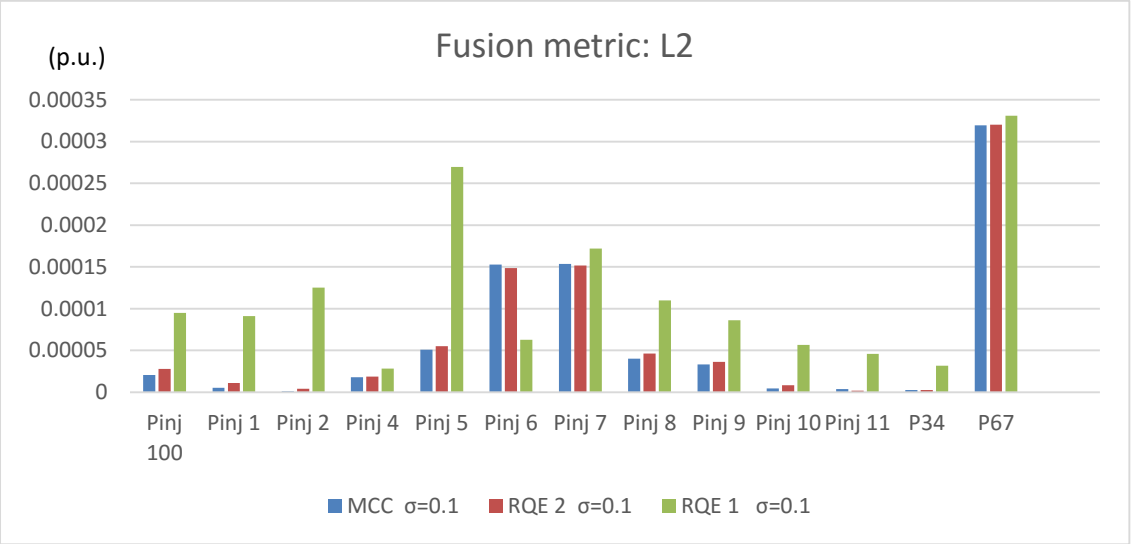

Figure B-3 - Residual errors in buses power injections and lines power flows using L2 as the fusion metric, with σ=0.1 and ignoring the residual error of the component Pinj3.

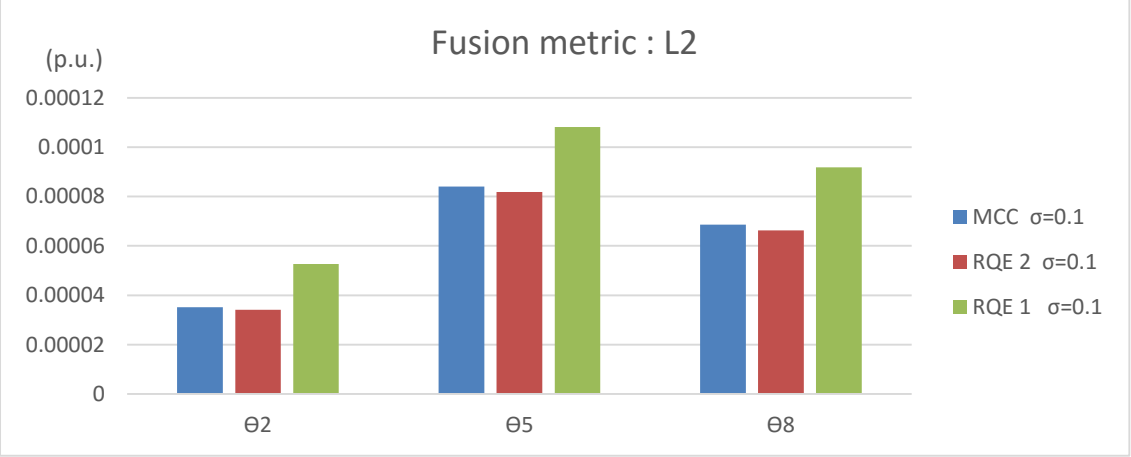

Figure B-4- Residual errors in voltage phases using L2 as the fusion metric, with σ=0.1.

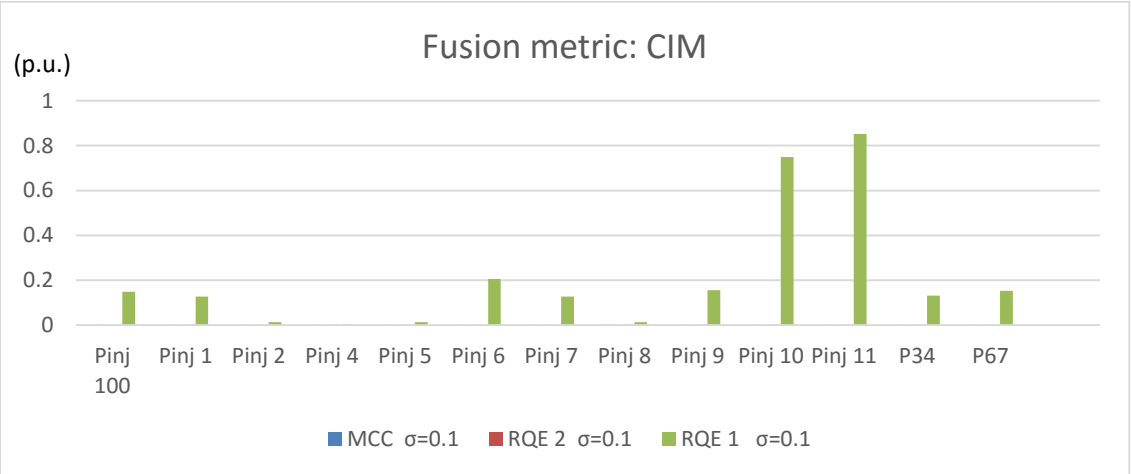

Figure B-5 - Residual errors in buses power injections and lines power flows using CIM as the fusion metric, with σ=1 and ignoring the residual error of the component Pinj3. Evaluation metrics with σ=0.1.

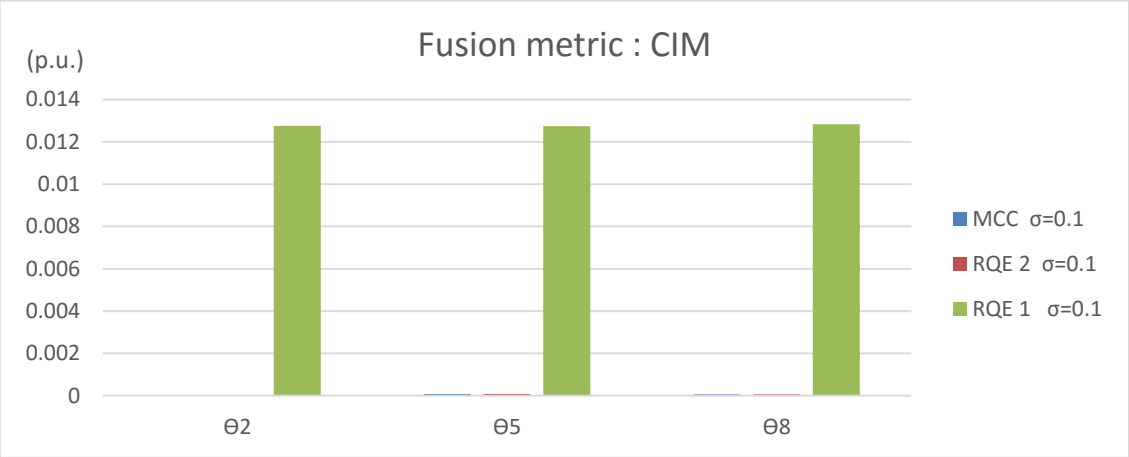

Figure B-6 - Residual errors for each estimation of voltage phases using CIM as the fusion metric, with σ=1. Evaluation metrics with σ=0.1.

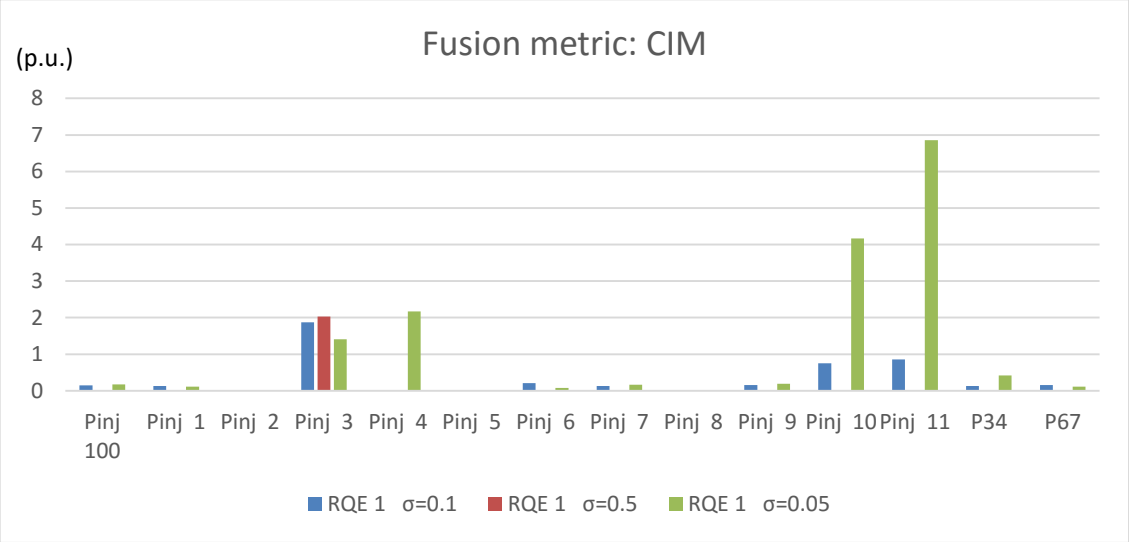

Figure B-7 - Residual errors in buses power injections and lines power flows, using CIM as the fusion metric with σ=1. Evaluation metric: RQE (case 1) with σ=0.1, σ=0.5 and σ=0.05.

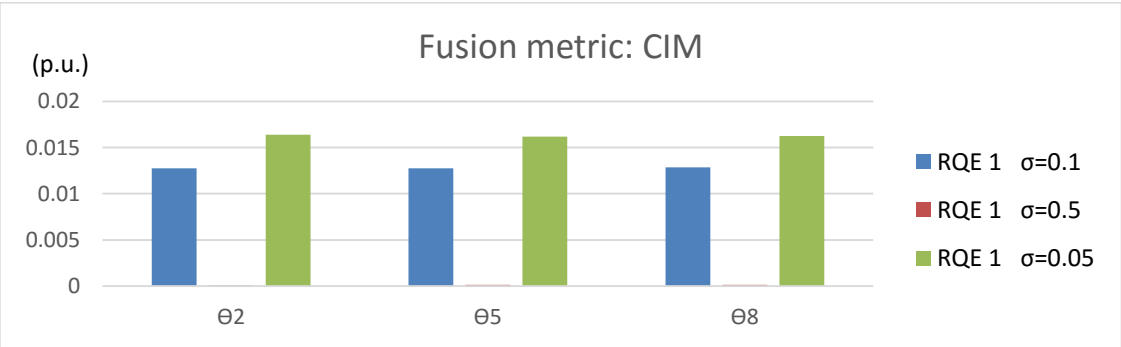

Figure B-8 - Residual errors in voltage phases, using CIM as the fusion metric with σ=1. Evaluation metric: RQE (case 1) with σ=0.1, σ=0.5 and σ=0.05.

Annex C

Article for submission

Considering the fusion method developed in this thesis, as well as the studies about state estimation concepts related with information theory, it was decided to write a paper. In this paper, the main ideas about a new sensory fusion perspective are presented. Also some experiences about the integration of concepts related with information theory in the developed fusion method are exposed. In this annex, a long abstract of this thesis is shown.

Sensory fusion applied to power system state estimation considering information theory concepts

Bruna Tavares[1], Vladimiro Miranda[2], Fellow, IEEE, Jorge Pereira[3]

***Abstract* –This letter proposes an innovative perspective about two topics of the state estimation process. The first is related to the fusion of two distinct sensory systems, where the attention paid to each of them can vary. The second aspect is about the replacement of the conventional optimization function (Weighted Least Square) by information theory related criterions, namely Correntropy and Renyi's Quadratic Entropy. These criteria aim to propose a novel way to identify and correct large errors.**

***Index Terms* — Correntropy, Information Theory, Renyi's quadratic entropy, Sensory fusion, State estimation.**

I. INTRODUCTION

THIS paper opens the discussion on a new perspective about the power system State Estimation (SE). The inclusion of PMU in the measurement system has been having a growing importance. However, the data fusion from conventional sensors and PMU has been proposed in a simple way: all the measurements are collected together with no distinction on which type of sensor are involved. In this paper an innovative vision about the data fusion in state estimation is presented.

Another aspect about the actual SE is its criterion. The Least Square Error (LSE) error is an optimal approach to estimate parameters only if the underlying distribution of errors is Gaussian. Thus, a gross error will cause disturbance in the estimation and, for that reason, should be identified and removed. Once LSE is not capable to ignore the gross errors, additional features are needed in the state estimator in order to filter possible outliers. The new paradigm proposed in [1] is based in maximizing the information that one can extract from the available measurements, minimizing the information on the residuals. In this work, it is extended to the sensory fusion context.

This paper presents a *proof of concept* in terms of a theoretical model and examples about a new perspective of sensory fusion and on how the adoption of information theory related concepts allow the natural identification and correction of gross errors, also in the sensory fusion context.

II. PARZEN WINDOW, RENYI'S QUDRATIC ENTROPY AND CORRENTROPY

The Parzen Window is a useful method to estimate a *pdf* distribution when the available data are discreet samples. It consists of applying in each sample a Gaussian kernel with a determined standard deviation [2]. The estimated distribution is given by:

$$\hat{f}_Y(z) = \frac{1}{n}\sum_1^n G(z - y_i, \sigma^2 I). \quad (1)$$

An ideal estimation would lead to a residual error distribution with no information, being represented by a Dirac function. If that distribution is centered at zero, it means that all errors are zero. In this way, a situation with minimum Entropy is found. One interested measure of entropy is Renyi's Quadratic Entropy (RQE) and is given by the following expression (2) [3]. Applying the PW method one gets (3).

$$H_2 = -log \sum_{i=1}^n p_{\varepsilon i}^2 , \quad (2)$$

$$H_2 = -log \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n G(\varepsilon_i - \varepsilon_i, \sigma^2 I) . \quad (3)$$

Correntropy is a measure of similarity between the two distributions [4]. In SE problem these distributions are estimated and measured values. Its expression is

$$\hat{V}(X,Y) = \frac{1}{n}\sum_{j=1}^n G(x_i - y_i, \sigma^2 I) = \frac{1}{n}\sum_{j=1}^n G(\varepsilon_i, \sigma^2 I), \quad (4)$$

where n is the size of the distributions, ε_i is the residuals for the component i and σ can be interpreted as the PW size.

III. SENSORY FUSION

The SE problem has the aim of minimize the estimation error. Once true values of the parameters are unknown, also the true value of the error is. What is available is the residual, the difference between the estimation and the measured values, given by $\varepsilon_i = z_i - \hat{z}_i$. The optimization of a function of the residual can be done through the minimization of its square error or RQE or by the maximization of the Correntropy, maximum Correntropy criterion (MCC). Let's assume an expression that evaluates a set of measures called fitness. The fitness evaluation can be done, among other options, by one of the three presented criteria, LSE, RQE or MCC.

In measurement systems with two sensory systems, it is possible to pay more attention to one system or to the other.

[1] e-mail: brunacostatavaresmail.com
[2] e-mail: vmiranda@inescporto.pt
[3] email: jpereira@inescporto.pt

This can be done assigning trust levels to the fitness function of each sensory system, fit_1 and fit_2. The fitness evaluation of the global system can be given by

$$fit = \alpha fit_1 + \beta fit_2, \quad (4)$$

where α represents the trust assigned to System 1 and β=1-α. The variation of the parameters α and β will influence the compromise reached. Thus, when α is 1 only System 1 is taken in consideration and its fitness evaluation reaches its maximum; at same time, the System 2 reaches its minimum. As α varies, the fitness varies gradually.

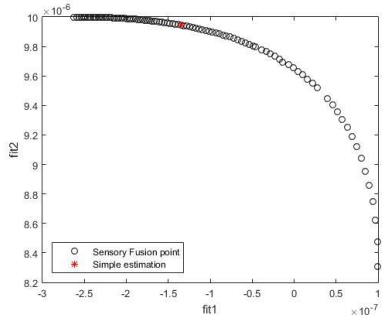

Figure 1- Pareto front obtained by α variation, represented by black circles. The red dot represent the point obtained by the simple fusion. The fitness function utilized was MCC, for the other metrics the form is equivalent.

As figure 1 evidences, there are a lot of non-dominated points different from the one obtained with the simple fusion. The question is: Which point is more close to the real state of the system? A way to obtain that point would by assigning stochastic values to the parameters α and β. To find those values, more studies are required, with research about the probability of each sensory system in leading to the real point.

Another way would be to choose the point closest to the *ideal*. The ideal point is the given by the coordinates (max fit_1, max fit_2). The metric found most convenient to measure the distance between both points was the absolute norm. L1 norm shows interesting properties: it's easily converted to the stochastic fusion and also its simplification doesn't require a pre-calculation of the point (max fit_1, max fit_2). The fusion objective functions for the SE are following presented, for LSE, MCC and RQE, respectively:

$$\min \ (\sum_{i\epsilon n1}(z_i - \hat{z}_i)^2 + \sum_{i\epsilon n2}(z_i - \hat{z}_i)^2), \quad (5)$$

$$\max \ (\sum_{i\epsilon n1} e^{-\frac{(z_i-\hat{z}_i)^2}{2\sigma^2}} + \sum_{i\epsilon n2} e^{-\frac{(z_i-\hat{z}_i)^2}{2\sigma^2}}), \quad (6)$$

$$\max(\sum_{i\epsilon n1}\sum_{j\epsilon n1} e^{-\frac{(\hat{\varepsilon}_i-\hat{\varepsilon}_j)^2}{2\sigma^2}} + \sum_{i\epsilon n2}\sum_{j\epsilon n2} e^{-\frac{(\hat{\varepsilon}_i-\hat{\varepsilon}_j)^2}{2\sigma^2}}). \quad (7)$$

IV. APPLICATION IN AC SYSTEM

A test has been conducted with a typical European medium voltage network, adapted from [5]. A measurement set was provided from a power flow solution and a gross error was introduced in bus 6 by eliminating its active power injection from 1.45 p.u. to 0 p.u.. The SE problem was solved with assistance of the EPSO algorithm. In this way, all the different objective functions (5), (6) and (7) were tested, with PW size 0.1. The residuals obtained, as well as their distribution, are exhibited in figure 2. While the LSE criterion led to high residuals in a large number of components, the MCC and RQE criteria provided an estimation with no significant residuals, except the residuals in $Pinj_6$, reaching 1.45 p.u.. That residual shows that MCC and RQE criteria were able to identify and correct the outlier, finding the true value of that measurement. The residual pdf for MCC and RQE approaches a Gaussian function centered at 0 and a peak on the value of the gross error. Thus, the outlier does not influence the estimation. The same is not true to LSE were the Gaussian gets distorted in the direction of the outlier, which means that the gross error not only was not identify but corrupted the estimation of other components.

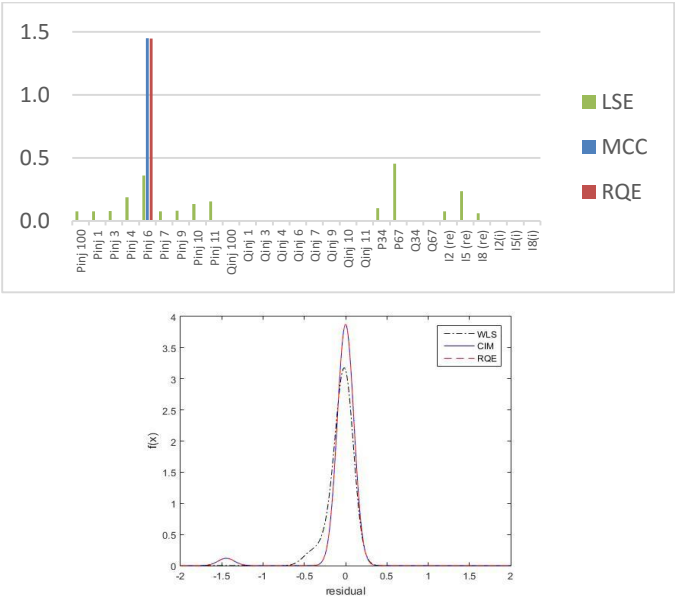

Figure 2- Top: abs. residuals (in p.u.) obtained by each method. The x-axis represents residual values in p.u.. Bottom: pdf of residuals obtained by LSE, MCC and RQE. The pdf was obtained with assistance of Parzen windows method with size σ = 0.1, Eq.(1).

V. CONCLUSION

This paper provides experimental results as *proof of concept* to the adoption of a new perspective in SE. First, it proposes a multi-criteria framework to achieve sensory fusion when in the presence of conventional and PMU devices. Second, it proposes to replace the traditional LSE by information theory based criteria, MCC and RQE – in this way the outliers are naturally isolated and ignored. The work shows that such adoption can be made, with positive results, in the context of sensory fusion. The challenge now is to develop an efficient algorithm for real-time use, taking advantage of these properties

REFERENCES

[1] V. Miranda, A. Santos, and J. Pereira, "State Estimation Based on Correntropy: A Proof of Concept," IEEE Trans. Power Syst., vol.24, no.4, pp.1888,1889, Nov. 2009.

[2] E. Parzen, "On the estimation of a probability density function and the mode", *Annals Math. Statistics*, v. 33, 1962, p. 1065.

[3] Principe, Jose C. *Information theoretic learning: Renyi's entropy and kernel perspectives*. Springer Science & Business Media, 2010. Chapter 2.

[4] Liu, W.F., P.P. Pokharel, and J.C. Principe, *Correntropy: A localized similarity measure.* 2006 IEEE International Joint Conference on Neural Network Proceedings, Vols 1-10, 2006: p. 4919-4924.

[5] Strunz, Kai, N. Hatziargyriou, and C. Andrieu. "Benchmark systems for network integration of renewable and distributed energy resources." *Cigre Task Force C*6 (2014): 04-02, pp33-37.