FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Information Theoretic State Estimation in Power Systems

Jakov Krstulović Opara

Doutoramento em Engenharia Eletrotécnica e de Computadores

Supervisor: Prof. Vladimiro Henrique Barrosa Pinto de Miranda, Full professor at FEUP

April, 2014

Information Theoretic State Estimation in Power Systems

Jakov Krstulović Opara

Doutoramento em Engenharia Eletrotécnica e de Computadores

April, 2014

Abstract

This thesis is a result of research work intended to shed some new light over the state estimation problem in power systems from perspective of information theory. The main motivation behind this investigation was enhancement of general robustness, reliability and flexibility in State Estimation process in respect to any type of inconsistencies that may appear in the input dataset. Moreover, all concepts presented in this thesis are built in a decentralized architecture in order to be compatible with modern power systems which are currently undergoing radical operational changes.

One of the main hypotheses of this thesis addresses the issue of adoption of paradigms from Information Theoretic Learning as novel state estimation criterion with multiple valuable properties. As a result of this work, a novel state estimation model that maximizes Correntropy as a robust localized measure of similarity between input values and their estimated values, is proposed. Substitution of the least squares criterion with the Maximum Correntropy Criterion (MCC) results in a fast and reliable estimator that naturally embodies all key bad data analysis functions. Moreover, it offers a strong theoretical justification for all previous attempts of modification of quadratic function in WLS estimator intended for establishment of robust properties. The novel MCC estimator is further improved via fusion with a mosaic of semi-autonomous auto-associative local measurement screening. Thanks to decentralized and localized architecture of the method, a high level of robustness is attained without affecting the real time feasibility for large scale systems.

The proposed information theoretic data screening is further extended to the problem of reconstruction of erroneous or incomplete information about system topology. The proposed competitive scheme of local autoencoders achieves remarkably efficient and robust reconstruction of entire topologies based on information backup between analog measurements and binary topology variables captured during the offline training process. Due to extreme robustness in respect to measurement redundancy and gross errors, the proposed topology estimator is particularly appropriate for state estimation in distribution power grids.

As a significant support to the proposed state estimation methodologies, a novel concept named Statistical Observability is established. Unlike the classical definition of observability, concerned with the analytical solution of a system of equations, the concept of Statistical Observability relies on the information content of joint probability distributions incorporated in the Maximum Mutual Information criterion. This concept allows for an autonomous definition of a decentralization degree and the most relevant local areas thus contributing to efficient topology estimation with minimal computational and memory requirements.

As the main contribution to theory of state estimation in power systems, this thesis paves the way towards new perspectives established on strong theoretical justification provided by Information Theory and Information Theoretic Learning. Moreover, it facilitates migration from global traditional concepts defined as problems of mathematical programming, towards decentralized local models that perceive state estimation as a problem of dealing with information.

Keywords: **power system state estimation, gross errors, information theoretic learning, Correntropy, autoencoders, artificial neural networks, power system topology, pattern recognition, observability, mutual information.**

Resumo

Esta tese é resultado de um trabalho de investigação destinado a lançar alguma luz sobre o problema de estimação de estado em sistemas de energia a partir da perspetiva da teoria da informação. A principal motivação por trás desta investigação foi o aumento da robustez geral, confiabilidade e flexibilidade no processo de Estimação de Estado em relação a qualquer tipo de inconsistências que podem aparecer no conjunto de dados de entrada. Além disso, todos os conceitos apresentados nesta tese são construídos em uma arquitetura descentralizada, a fim de ser compatível com os sistemas de energia modernos, que estão atualmente passando por mudanças operacionais radicais.

Uma das principais hipóteses desta tese aborda a questão da adoção de paradigmas de Aprendizagem com base em Teoria da Informação como novo critério de Estimação de Estado com várias propriedades valiosas. Como resultado deste trabalho, propõe-se um modelo de Estimação de Estado que maximiza a Correntropia como uma medida localizada robusta de similaridade entre os valores de entrada e os valores estimados. A substituição dos mínimos quadrados pelo critério da máxima Correntropia (MCC) resulta em um estimador rápido e confiável que naturalmente incorpora todas as funções fundamentais de análise de erros nos dados. Além disso, oferece uma justificação teórica forte para todas as tentativas anteriores de modificação da função quadrática no estimador WLS (Mínimos Quadrados Pesados), visando o estabelecimento de propriedades robustas. O novo estimador MCC é melhorado através do tratamento de medidas por integração num mosaico de redes locais autoassociativas. Graças à arquitetura descentralizada e localizada do modelo, um elevado nível de robustez é alcançado sem afetar a viabilidade de tempo real para os sistemas de grande escala.

O processamento dos dados com base em Teoria da Informação é alargado para o problema da reconstrução de informações erradas ou incompletas sobre a topologia do sistema. O esquema competitivo proposto de autoencoders locais alcança uma reconstrução extremamente eficiente e robusta da topologia de um sistema, com base no conteúdo de informação partilhado pelas medições analógicas e variáveis de topologia binárias, que fora capturado durante o processo de treino offline. Devido à elevada robustez em relação à redundância de medição e erros grosseiros, o estimador de topologia proposto é particularmente apropriado para Estimação de Estado em redes de distribuição de energia.

Como um apoio significativo para as metodologias de Estimação de Estado propostas, um novo conceito chamado de Observabilidade Estatística é estabelecido nesta tese. Ao contrário da definição clássica de observabilidade, preocupado com a solução analítica de um sistema de equações, o conceito de Observabilidade Estatística se baseia numa medida do conteúdo de informação da distribuição de probabilidade conjunta traduzida pelo critério de Máxima Informação Mútua. Este conceito permite uma definição autónoma de um grau de descentralização e das áreas locais mais relevantes, contribuindo assim para uma estimação de topologia eficiente, com requisitos mínimos de poder de cálculo e memória.

Sendo uma contribuição forte para a teoria da estimação de estado em sistemas de energia,

esta tese abre o caminho para novas perspetivas, estabelecidas sob forte justificação teórica fornecida pela Teoria da Informação. Além disso, facilita a mudança de paradigma a partir de conceitos tradicionais globais, definidos como problemas de programação matemática, para modelos locais descentralizados que apercebem a estimação de estado como um problema com base em informação.

Acknowledgments

This thesis was made possible through selfless support of a number of great people.

First of all, I would like to thank my supervisor Prof. Vladimiro Miranda for finding the time in his busy schedule to help and guide me throughout the course of this thesis. His encouragement and support were key factors in my decision to start my Ph.D. at FEUP. Thanks to his openness to the international research and industrial community, I have had the opportunity to collaborate with some outstanding researchers and engineers. I would also like to thank him for respecting my decision to return to my country and his patience and understanding during our long-distance correspondence. Finally, I would like to thank him for being a continuous source of inspiration and promising ideas.

Next, I would like to thank Prof. Jorge Pereira, for accepting the role of my co-supervisor. His invaluable technical advice has considerably improved my progress in the early beginnings of my research work at INESC Porto.

I would also like to express my gratitude to Prof. Antonio Simões Costa for his collaboration on our research project. The experience of working with Prof. Simoes Costa has proven to be priceless in the context of my understanding of the research area addressed in this thesis.

Next, I would like to thank Prof. Antonio Gómez Expósito, Prof. Antonio Simões Costa, Prof. Manuel Matos, Prof. Fernando Barbosa, António Carrapatoso and Prof. Jorge Pereira for agreeing to be members of the PhD thesis defense committee.

A special thanks goes to my great friend and colleague Hrvoje Keko, for his priceless help and support during my application process for a junior-researcher position at INESC Porto. Also, for his great help during my process of acclimatization to a new life, language and work environment in Porto.

For making my life in Porto especially joyful, I must thank my outstanding friends: Hamed Akhavan, Arif Ur-Rahman, Julija Vasiljevska, Stanislav Stoykov, Danica Drpić and Rosa Crespo. Countless lunches, dinners, coffees and unforgettable trips in their company have helped me take my mind off research and have left me now with a strong feeling of *Saudades*.

I would also like to thank to Prof. Ranko Goić for giving me the opportunity to work in my hometown. His contributions in early beginnings of my engineering and research career were of key importance for my further progression and orientation. His understanding and encouragement have played an important part in my decision to apply for the research project at INESC Porto and the subsequent Ph.D. at FEUP.

I would like to thank to Dinko Relković for accepting the tremendous work of reviewing my English.

My open-space-mates from INESC Porto were a joy to be around, and always prompt to help in any regard.

A special thanks goes to my closest colleagues from University of Split: Damir Jakus, Ivan Penović and Josip Vasilj. First of all I would like to thank them for being great friends but also for being the best colleagues.

A special thanks to dearest Martina and Marin Barišić, my Portuguese-Croatian adopted family. I am not only extremely grateful to Martina and Marin for accepting me every time as a couch-surfer but also for the immense joy they brought in my life spent in Porto.

My graduate studies abroad would barely have been possible without the unconditional love and endless support from my family. First of all, I owe my sincerest gratitude to my mother Ana and father Vlado, for allowing me to choose my own path and for encouraging me along the way. I am also grateful to my two incredible grandmothers Drava and Rada, whom I consider to be my endless sources of inspiration and life wisdom.

Finally, the most special place in the acknowledgments belongs to Danijela. The incredible amount of patience and understanding she has shown me during my PhD work can only be explained with true love. Close to me, during each step of the way, she was always there as my best friend, a psychotherapist and a source of good mood. If I could dedicate this thesis to one person alone, it would definitely be my wife, Danijela.

The work reported in this thesis, was supported in part by the ERDF from the EU through the Programme COMPETE and by the Portuguese Government through FCT — Foundation for Science and Technology, project ref. GEMS PTDC/EEA-EEL/105261/2008 and LASCA PTDC/EEA-EEL/104278/2008.

Jakov Krstulović Opara

Contents

CONTENTS

List of Figures

List of Tables

Abbreviations

AANN	Auto-associative Artificial Neural Network (Autoencoder)
ANN	Artificial Neural Network
CDF	Cumulative Density Function
CIM	Correntropy Induced Metric
DAANN	Denoising Auto-Associative Neural Networks (Denoising Autoencoders)
DAANN-EPSO	Data screening with denoising autoencoders and EPSO
DAANN-MCC	Denoising Autoencoder-Aided MCC power system state estimation
DMS	Distribution Management System
DSE	Distribution State Estimation
DWLS	Dependent Weighted Least Squares
EEC	Error Entropy Criterion
EMS	Energy Management System
EPSO	Evolutionary Particle Swarm Optimization
FASE	Forecasting-Aided State Estimation
GE	Gross Error
GMCC	Generalized Maximum Correntropy Criterion
GSE	Generalized State Estimation
HTI	Hypothesis Testing Identification
IP	Information Potential or Quadratic Information Potential Estimator
IT	Information Theory
ITL	Information Theoretic Learning
LAV	Least Absolute Value
LMR	Least Measurements Rejected
LMS	Least Median of Squares
LNRT	Largest Normalized Residual Test
LRT	Largest Residual Test
LS	Least Squares
LTS	Least Trimmed Squares
MCC	Maximum Correntropy Criterion
MCC^H	Norma-scaled Maximum Correntropy Criterion\
MCC^N	Normalized Maximum Correntropy Criterion
MCIM	Minimum Correntropy Induced Metric
MCS	Maximum Constraints Satisfaction

MEE	Minimum Error Entropy
MI	Mutual Information
MIFS	Mutual Information Feature Selector
MINLP	Mixed Integer Non-Linear Problem
MMI	Maximum Mutual Information
MMSE	Minimum Mean Square Error
MNMR	Maximum Normal Measurement Rate
MSE	Mean Square Error
NLP	Non-Linear Problem
NQC	Non-Quadratic Criterion
OPF	Optimal Power Flow
PCA	Principal Component Analysis
PDF	Probability Density Function
PMU	Phasor Measurement Unit
POCS	Projection Onto Convex Sets
PSSE	Power System State Estimation
QC	Quadratic-Constant Criterion
QL	Quadratic-Linear Criterion
QR	Quadratic-square Root Criterion
QT	Quadratic-Tangent Criterion
SE	State Estimation
SO	Statistical Observability
TC	Topology Cell
TSE	Topology State Estimation
WLAV	Weighted Least Absolute Value
WLS	Weighted Least Squares
w/	with
w/o	without

List of symbols

State estimation symbols

x	state vector
x^k	state vector at iteration k
Δx^k	iterative step size of the state vector at iteration k
$\hat{x}$	estimated state vector
x^{true}	true state vector
$J(x)$	scalar objective function of errors
$J_{MCC}(x)$	scalar objective function using Maximum Correntropy Criterion
$J_{WLS}(x)$	scalar objective function using Weighted Least Squares criterion
$c(x)$	equality-constraint functional vector
$f(x)$	inequality-constraint vector
z	measurement vector
z^{true}	true measurement vector (load flow result)
$h(x)$	non-linear function vector that relates vector z to the state vector x
e	measurement error vector
r	residual vector
g	gradient vector
z_i	i^{th} measurement
$h_i(x)$	non-linear function that relates i^{th} measurement to the state vector x
e_i	i^{th} measurement error
r_i	i^{th} residual
m	number of available measurements
m^r	number of available measurements
n	dimensions of the state vector x, i.e. number of state variables
R	covariance matrix of measurement errors
H	Jacobian matrix
G	gain matrix
Ω	covariance matrix of measurement residuals
T	variable which defines topology of a topology cell
P_i	active power injection measurement in bus i
Q_i	reactive power injection measurement in bus i

P_{i-j}	active power flow measurement through line $i-j$
Q_{i-j}	reactive power flow measurement through line $i-j$
V_i	voltage magnitude measurement in bus i
θ_i	voltage angle at bus i
$\varepsilon^V_{abs,\omega}$	average absolute error of the voltage-magnitude estimates regarding the true values for scenario ω
$\varepsilon^\theta_{abs,\omega}$	average absolute error of the voltage-angle estimates regarding the true values for scenario ω
t_ω	state estimation computational time for scenario ω

Information Theoretic symbols

$H(X)$	Shannon entropy
$H_{R\alpha}(X)$	Renyi's entropy of order α
$H_{R2}(X)$	Renyi's quadratic entropy
$\hat{V}(X)$	quadratic information potential estimator
$D_\alpha(p\|\|q)$	Renyi's divergence of order α
$I(X,Y)$	Shannon mutual information
$I_\alpha(X,Y)$	Renyi's mutual information of order α
$I_2(X,Y)$	Renyi's quadratic mutual information
$p(x)$	probability density function
σ	standard deviation of the Gaussian kernel
$G_\sigma(\mu)$	Gaussian kernel with mean μ and standard deviation σ
$\upsilon_\sigma(X,Y)$	Correntropy with kernel bandwidth σ
$\hat{\upsilon}_\sigma(X,Y)$	Correntropy estimate with kernel bandwidth σ

Chapter 1

Introduction

1.1 Motivation

Any control decision in an electric power system, whether automatic or manual, relies on the access in real time to the actual state of the system. As the state is never directly available, there is a process called state estimation that estimates the state variables and derives their values from measurements that are being periodically acquired by the SCADA. Hence, the problem is traditionally approached by solving a non-linear system of equations that characterizes an operational point. The accuracy and robustness that need to be achieved in real time are of key importance for such a process, since the knowledge about the true system is a prerequisite for the efficiency of an Energy Management System (EMS) function, while any bias from the true state may result in inappropriate or even dangerous control decisions. In spite of decades of reasonably successful efforts to develop a robust method that may reliably be used by the industry in large scale systems, the problem of Power Systems State Estimation (PSSE or simply SE) still remains in the focus of researchers' attention.

During the years of practice in transmission systems, a standard practice has been accepted in PSSE: regression models based on Weighted Least Squares (WLS) criterion (a variant for the minimum Mean Square Error criterion or MSE) and the Newton algorithm as a solver. In vertically integrated power systems, high levels of monitoring, automation and remote control are traditionally limited to transmission grids, while distribution grids have not been in the focus of operational effectiveness. This approach is acceptable whenever the distribution network is considered as quasi-static by its nature, i.e. with well-understood and predictable load and voltage profiles and with no need for reconfiguration or alteration of the protection and control settings under some normal circumstances.

The modern power system concepts are influenced by deregulated electricity market conditions and the opening of power systems to non-utility players. Accordingly, enabled by the new regulatory frameworks, some fundamental operational changes are induced. These new circumstances particularly affect distribution networks which are beginning to face an exploding installation of distributed generation, together with the ever-demanding requirements for quality

of service and operational efficiency. It is a migration from a static radial configuration with a one-way energy flow to a system that supports distributed resources and some new types of loads (thinking about actual trends for large scale implementation of electric vehicle usage).

This type of network will undoubtedly require more sophisticated control and automation, which is barely implementable at the present time, primarily due to a very poor monitoring infrastructure. Therefore, in order to cope with the new challenges, large investments are being directed primarily towards the extension of system monitoring down to the full extent of distribution grids. Such an abundance of new data flow needs to be dealt with by a Distribution Management System (DMS) that will, through sophisticated automation applications, meet the new requirements of distribution system functionality.

The grid is growingly receiving a large variety of distributed generation and even the energy storage is appearing, especially when organized local stand-alone entities called micro-grids are allowed to exist. In this context, with added sources of uncertainty (in load, renewable generation or car batteries), a decentralized management process seems to be a logical solution; to process such an extensive amount of data and to control such a stochastically behaving system. This converges with a vision for future power systems, popularly addressed with a term smart grids. Advanced management and control technologies which use information and communications technology are now required, with a principal aim to improve the efficiency, sustainability and quality of service of the electric energy production and distribution. Apart from giving support to the operation of distribution grids with active sources (such as renewable generation), these new concepts are expected to bring benefits for both utilities and their end customers.

The state estimator, as a core function of EMS and DMS, revives again the interest of researchers in order to keep it in line with the current trends. Therefore, in the new deregulated and technically evolved power system environment, the state estimation function will also require to be upgraded with some additional features in order to meet the following emerging issues:

- ▷ the expansion of SE to distribution networks with high uncertainties and difficulties in observability,
- ▷ modern distribution grid concepts – micro-grids and smart grid,
- ▷ progressive introduction of Phasor Measurement Units (PMU),
- ▷ increasing requirements of accuracy and robustness together with heavier demands of computing power, induced by shorter estimation cycles for continuously growing massive data bulks.

The traditional, well-established state estimation procedure solves a non-linear problem based on the measured inputs of active and reactive power, voltage and current magnitudes, as well as the reported binary variables about breaker statuses. The classical objective of the estimation process has been set to minimize the variance of measurement residuals (difference between the measured and the estimated values). However, with the advent of new operational circumstances,

the state estimation concept and procedures need to be adapted to multiple perspectives and strongly enhanced. This is the fundamental motivation for the work presented in this thesis.

1.2 Specific Challenges

In power system state estimation, one of recently most investigated issues is the implementation of a rapidly growing number of PMUs, which supply, with high frequency, time-stamped voltage or current angle and magnitude measurements. Given that PMUs directly measure the state vector, one can ask about the necessity for a state estimator in the near future. However, the ever present noise in measurement readouts inevitably demonstrates some degree of error. Therefore, in order to estimate the state of a system, some level of redundancy should always be expected from obtained measurements. Unlike the classical non-linear problem, a state estimator based on pure PMU measurements is linear. However, in the near future and perhaps for some time, a transitional need for a hybrid model that will process in parallel both PMU and SCADA measurements is foreseen.

Another issue is the Distribution State Estimator (DSE), which is becoming necessary with the growth of active sources in distribution networks. The centralized automation and control of distribution systems will therefore no longer be a dominant paradigm. One must be aware that full monitoring, on the level that exists in a transmission system, is not realistically expected at the distribution level for obvious economic reasons. Therefore, a compromise solution needs to be found somewhere among the selective deployment of low-cost sensors, smart metering information and an appropriate DSE. In such an environment, the input dataset that needs to be conveyed to a state estimator is, first of all, enormous in size, and it is still prone to provide limited observability and questionable accuracy as long as redundancy remains small and sensor interrogation becomes a huge bottleneck. Due to all this, a future distribution estimator will significantly differ in some ways from the concept that is traditional in transmission systems. An envisageable solution is to decentralize and distribute the process to logical and functional local areas. However, because of the requirements for redundancy and observability, the generation of coherent pseudo-measurements will become a key issue.

The modern way of organizing the electric power networks, under the concepts of micro-grids and smart grids, assumes some advanced functions of automation and control, with a general aim to improve efficiency and reliability of electrical supply and the conditions for high penetration of stochastic (renewable) energy sources and electric vehicles. For the establishment of smart grid technologies, a tremendous amount of real-time and operational data is required, with an increase in the number of sensors and the need for more information on the system operation. Therefore, the function which is the main supplier of information about the system operation is exactly the state estimator. With such an increased requirement on monitoring and control of the distribution system, the demand for accuracy, reliability and robustness of the state estimator becomes more stringent. Therefore, it is also desirable that a modern (distribution) state estimator possesses some smart features in order to meet the new challenges and provide functionalities for building modern electric power systems.

The extension of SE to distribution networks brings to light the importance of estimation of topology. Distribution networks are much less stable in topology than transmission systems: switches and breakers may be operated for a number of reasons, from isolation of failures to reduction of losses, improvement of security and reliability or even to load transfer between substations or between feeders. The breaker status signals are binary by nature – and an erroneous signal (possibly from a failure in a sensor contact) is necessarily a gross error with a profound impact on the SE result. In addition, it is relatively common in distribution systems that breaker status signals are missing (not received at the SCADA for transmission reasons), which still must be classified as a gross error. Therefore, the estimation of topology is a big challenge for the future state estimation procedure.

Finally, in conceptual terms all the classical SE procedures, namely the ones in current use in the industry, have a flaw: their estimation of gross errors (or even non-gross errors) originating from provisional estimation is itself already contaminated by the errors. In other words, the main trend is development of models that explore in one way or the other the importance of the residuals, i.e. the difference between the measurement values and estimated values, the calculation of which ignores the nature of the noise present in the data. This GE impact may generate what is known as masking effect, which represent the main burden for reliability of most of the robust estimators. The residuals are not, in the present SE paradigm, true errors, i.e., they are not a difference between measurements (with errors) and true values. This results from a two-step procedure: first a provisional estimation and then the attempt to extract conclusions from the residuals. This means that the SE procedure takes, as a reference, not the true value or true data pattern but an assumed and approximate pattern.

The challenge is therefore to develop a new approach that takes as reference the (hidden) pattern of true values, i.e., the true manifold supporting the power flow equation solutions. In this case, gross errors will be seen as outliers from the true pattern and will therefore be prone to elimination and replacement by true values lying on the true manifold. Conceptually, this is a major change from the traditional point of view, and a success in this development will provide a sound theoretical justification for a number of novel models and procedures that will embody the new SE paradigm.

1.3 Thesis Objectives

The state estimation in a power system could be an easily solvable problem if all the real-time data, provided by SCADA system, could be taken as correct. In that ideal case, when all the analog and binary values are electrically coherent with one another and in accordance with the true state, only with a very small noise-signal ratio, the most suitable practical solution is to minimize the sum of squared measurement residuals. That is the well-known Weighted Least Squares (WLS) state estimator, which is definitely the best estimator under the assumption that the measurement errors follow the Gaussian distribution. In addition, thanks to its suitability for online use, the WLS is nowadays still a widely accepted industrial solution.

Unfortunately, the SCADA is prone to also convey corrupted data or simply it does not assure the simultaneous availability of all measurements at all times. In that case, the WLS does not have a tool to recognize such bad data, and it provides a state of the system that can be significantly biased. Hence, an inevitable component of any state estimator is the bad data processor. In that regard, a large spectrum of different approaches was developed for detection, identification, elimination or even correction of the gross errors. So far, the problem has not received a unique wide-accepted solution, able to reveal all possible bad data in real time, especially for large scale systems. The difficulties orbit around the trade-off between robustness and computational requirements of the system.

When talking about inconsistencies the data provided by SCADA can contain, they are not just limited to analogue measurements of power or voltage, but also to binary variables that define breaker statuses, i.e. system topology. When topology information is forwarded to a state estimator, such as in the case when some of the breaker statuses are wrong, it can significantly corrupt the estimated state, and then possibly result in an inappropriate control action. Therefore, some special tools are required for detection of system configuration that is not in accordance with other available electric values.

This thesis intends to address the challenges identified above and open a new perspective on the state estimation problem, funded on the principles of information theory. The main research motivation comes from the above explained space for improvement that exists in the state of the art state estimators. The intention is to also cope with challenges that will emerge with radical operational changes that nowadays power systems are undergoing. The target focus is on enhancement of robustness regarding the gross errors and signals that are missing. The result is a novel state estimator paradigm that offers an appropriate robust theoretical justification, which can generally be applied to any type of data. Moreover, thanks to the properties of information theoretic descriptors that the proposed model is based on, this thesis aims to propose a concept suitable for application within a modern concept of decentralized/distributed system control.

Hereafter, the main objectives of the research reported within this thesis is summarized:

- A new state estimation model should aim at extracting the maximum information possible from the conjunction of data set (measurements) and the global knowledge of the system properties (power flow equation supporting manifold). Information quantity evaluation should be possible.

- The proposed state estimator should be robust to any type of bad data and at the same time be feasible in the real-time environment.

- It should be able to detect, identify and quantify gross errors and then provide an accurate state of the system that is uncorrupted, even in low redundancy circumstances, or when the bad data are critical.

- The model should be able to reconstruct electrical signals that are missing, i.e. that are suddenly not conveyed by the SCADA system. This reconstruction should be based on information that exists in the known values and on data stored in the historical database.
- A specific tool should be developed for processing topology variables. The objective should be to develop a robust topology estimator, which should detect and identify inconsistencies among the provided breaker statuses and afterwards reconstruct the true values based on the other known electric variables.
- The model should adopt a decentralized and distributed architecture in order to be applicable to small or large scale problems through a natural scalability process of the replication of estimation cells of roughly constant size, instead of a growth in size of a single cell.
- The application of information theory concepts should be the unifying aspect of the new SE procedures, and should open the way to solving problems on the relative importance of distinct measurements and on the optimization of data collection.
- The new SE approach should produce results of at least comparable quality for trivial cases where the current paradigm (WLS) performs well and yet display net advantages when dealing with specific gross error cases, including the topology estimation.

Generally, the intention is to establish a new paradigm that will offer a strong theoretical justification and open way to some new perspectives in state estimation theory.

1.4 Structure of the Thesis

This thesis is organized in seven chapters. Regarding the analyzed issues of the state estimation, it may be divided in two main parts. The first part (Chapters 3 and 4) deals with gross errors in (analogue) measurements, and the second part (Chapters 5 and 6) discusses estimation and reconstruction of a power system topology.

Chapter 1 introduces problems and presents the most recent challenges in power system state estimation, as well as motivations and objectives of the research work which is presented in the subsequent 6 chapters and 4 appendices.

In Chapter 2 of this thesis, an extensive overview of the existing state estimation models is presented, with a special emphasis on robust state estimators, particularly referring to static and dynamic (forecasting-aided) approaches. It does not just refer to state of the art, but gives a summarized overview of the historical evolution of robust state estimation, from early beginnings to modern times. Another key concern of this chapter are achievements in topology processing of a power system. Finally, a particular emphasis is put on the state of the art and the most actual challenges the state estimators are expected to face within the modern power system concepts.

In Chapter 3, a new paradigm for power system state estimation from the information theoretic point of view is established. The novel estimation objective is set to maximum Correntropy

criterion, as a measure of similarity between measured and estimated values. The Newton's method is adopted as an appropriate solver for the real-time environment and is coupled with the proposed adaptive Parzen annealing strategy. Via the various case studies and comparative analysis with some conventional estimators, all the specific properties, benefits and limitations of the proposed estimator are demonstrated, mostly regarding the robustness to gross errors and computational requirements.

In Chapter 4 the Auto-Associative Neural Networks (autoencoders) are proposed for data pre-screening, before they are submitted as an input to a state estimator. A decentralized and localized method is established to perform reconstruction of missing data and detection, identification and quantification of gross errors. Finally, the method is coupled with the Correntropy estimator from Chapter 3, resulting in an hybrid industry-acceptable state estimation method, characterized by high robustness in any gross error scenario.

Chapter 5 focuses on reconstruction of unknown or missing information about the power system topology. The Auto-Associative Neural Networks are adopted for extraction of information about the system topology from the other available measured electric values. A decentralized topology estimation is proposed, based on competitive architecture of local autoencoders. This concept is especially suitable for distribution networks, i.e. networks with poor measurements redundancy and frequent changes in topology configuration.

Chapter 6 offers an additional theoretical support to the concepts proposed in the previous chapters. The Mutual Information is proposed as a criterion for an automated definition of more efficient decentralized topology estimator from Chapter 5. The concept is then extended to a general measure of information among any type of variables, establishing a novel paradigm called Statistical Observability.

Chapter 7 presents a brief summary of the thesis and a discussion about possible future research work.

This Thesis ends with two Appendices. Appendix A provides the network test cases that are used throughout this research.

Appendices B and C.2 provide theoretical background of the concepts adopted in this thesis.

Appendix D lists all the papers published as a result of the research work presented and reported in this thesis.

Chapter 2

State of the Art - Problems Overview

This section aims to present some of the mile-stones in theory of state estimation in power systems. In addition, it also provides a relatively exhaustive literature overview including not only the state-of-the-art work, but also an overall review of the relevant literature and methods established in the theory of power system state estimation. This is considered to be inevitable and of key importance since the methods which are dating from the early 70's, the time when state estimation was established, are still actual nowadays.

2.1 State Estimation in Power Systems - the Traditional Model

Generally speaking, the state estimation is a process of determining the most likely state of a system based on the measured quantities. Accordingly, any state estimator can be formulated as a following optimization problem:

$$\min_{x} \; J(x) \tag{2.1}$$

subject to

$$c(x) = 0$$
$$f(x) \leq 0$$

where:

- x vector of state variables,
- $J(x)$ scalar function of errors,
- $c(x)$ equality-constraint functional vector,
- $f(x)$ inequality-constraint vector.

Definition of the objective function $J(x)$ depends on the criterion used for solving the state estimation problem.

Estimation of the state in a power system was first proposed by Schweppe *et al.* in [1]–[3] as a data processing algorithm which obtains an estimate of the static-state vector based on redundant

meter readings and other available system information. The problem is formulated in order to find values of a set of state variables in the state vector x that satisfy the following equations:

$$z_i = h_i(x) + e_i, \quad i = \{1, ..., m\} \tag{2.2}$$

where:

- x state vector of dimension n,
- z_i i^{th} measurement,
- h_i non-linear function that relates i^{th} measurement to the state vector x,
- e_i error of the i^{th} measurement,
- m number of available measurements.

Equation (2.2) includes m available measurement values in vector z and a set of non-linear relations $h(x)$ of n state variables in the state vector x with measurements z and errors of measured values in vector e. In fact, it is an overdetermined system of nonlinear equations, with m measurements and n state variables, where $n < m$. The result of this state estimation process will be estimative of the state vector $\hat{x}$. The difference between estimated measurements vector $\hat{z} = h(\hat{x})$ and measured values z will give an estimate of the errors, i.e. residuals r:

$$r = \hat{z} - z = h(\hat{x}) - z \tag{2.3}$$

The real metering errors can be expressed as a difference between the measured value and the unknown true value for the true state vector: $e_i = z_i - z_i^{true} = z_i - h_i(x^{true})$. The true state vector x^{true} is a theoretical value that is never known in practice. Thus, real metering errors e are also unknown, but only their estimate, i.e. residual vector r.

In relation to the metering errors, the following assumptions are traditionally made about their statistical properties:

- ▷ $E(e_i) = 0, \quad i = 1, \ldots, m$ (expected, mean value is equal to zero),
- ▷ $E[e_i e_j] = 0$ (measurement errors are assumed to be independent),
- ▷ $Cov(e) = E[e \cdot e^T] = R = diag\{\sigma_1^2, \ldots, \sigma_m^2\}$, where σ_i represents a standard deviation and R covariance matrix.

The key important issue is to define the scalar objective function $J(x)$ from equation (2.1). The first power system state estimator was proposed in [1] as an optimization procedure that minimizes weighted sum of the residuals squares. Put simply, they adopted a sum of squared residuals for $J(x)$ and established a ***Weighted Least Squares*** (**WLS**) power system state estimator, which is nowadays still the most used estimator in practice. Hence, the objective function of the WLS state estimator is defined as the following:

$$J(x) = \sum_{i=1}^{m} \left(\frac{z_i - h_i(x)}{R_{ii}} \right)^2 = \sum_{i=1}^{m} \left(\frac{r_i}{R_{ii}} \right)^2 = [z - h(x)]^T R^{-1} [z - h(x)] \tag{2.4}$$

subject to

$$c(x) = 0$$
$$f(x) \leq 0$$

where:

- r vector of measurement residuals,
- R covariance matrix of measurements errors,
- $c(x)$ equality-constraints vector which represents functional vector that enforces conditions at zero-injection buses (no generation and no demand),
- $f(x)$ inequality-constraints vector which models physical limits of the system.

Since measurement errors are assumed to be independent, covariance matrix R is diagonal with standard deviations σ_i for each particular measurement. The definition of these variances is usually based on the accuracy of corresponding measurement devices. Thus, matrix R directly tunes the impact of each measurement on the final estimation result. In literature, an expression of covariance matrix can be found through the weighting matrix $W = R^{-1} = diag\{1/\sigma_1^2, \ldots, 1/\sigma_m^2\}$ which weights measurements inversely proportional to their variances.

The optimization problem from eq. (2.4) is traditionally being solved by Newton-Raphson method. This numerical method is applied in order to meet the first-order optimality conditions [4]:

$$g(x) = \frac{\partial J(x)}{\partial x} = -\sum_{i}^{m} \frac{z_i - h_i(x)}{R_{ii}} \nabla h_i(x) = -H^T(x) R^{-1} [z - h(x)] = 0 \tag{2.5}$$

where $H(x)$ is the Jacobian matrix of $m \times n$ dimensions:

$$H(x) = \begin{bmatrix} \frac{\partial h_1(x)}{\partial x_1} & \cdots & \frac{\partial h_1(x)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m(x)}{\partial x_1} & \cdots & \frac{\partial h_m(x)}{\partial x_n} \end{bmatrix} \tag{2.6}$$

Expanding the non-linear function $g(x)$ into its Taylor series around the state vector x^k gives the following result:

$$g(x) = g(x^k) + G(x^k)(x - x^k) + \ldots = 0 \tag{2.7}$$

After neglecting the higher order terms, an iterative solution scheme follows:

$$G(x^k)\Delta x^{k+1} = g(x^k) = \nabla J(x^k) = H^T(x^k) R^{-1} [z - h(x^k)] \tag{2.8}$$
$$x^{k+1} = x^k + \Delta x^{k+1} \tag{2.9}$$

where x^k is the state vector at iteration k and $G(x)$ is a sparse, positive-definite symmetrical matrix called *Gain matrix*. It is calculated as a second derivative of the objective function:

$$G(x) = \nabla^2 J(x) = \frac{\partial g(x)}{\partial x} = H^T(x) \cdot R^{-1} \cdot H(x) + \sum_i^m \frac{z_i - h_i(x)}{R_{ii}} \nabla^2 h(x) \tag{2.10}$$

The second term in equation (2.10) represents second-order information of $\nabla^2 J(x)$, which is mainly neglected in practice in order to avoid an additional calculation of the $mn \times n$ Hessian matrix $\nabla^2 h(x)$. This simplification reduces classical Newton-Raphson method to the so-called Gauss-Newton method. Therefore, the gain matrix for the Gauss-Newton method (also called the Gauss-Newton Hessian) is defined by:

$$G(x) = H^T(x) \cdot R^{-1} \cdot H(x) \tag{2.11}$$

The justification of this simplification, for state estimation in power systems, was exhaustively investigated by van Amerongen in [5]. Amerongen showed that simplification within Gauss-Newton method is justified in most of the cases and is even preferable in the first few iterations. Only in some extreme cases with specific gross errors scenarios, the convergence of Gauss-Newton method can be threatened and a shift to Newton method is required. Finally, equations (2.8) and (2.11) represent an iterative procedure for WLS state estimation. What remains to be defined is the initial guess for the state vector x^0 which is usually assumed to the flat voltage profile, i.e. all the bus voltages are assumed to $1.0 p.u.$ and in phase with each other ($\theta_i = 0^\circ, \forall i = 1, ..., n$). In vast majority of scenarios, the flat start is within the convenient convex zone which assures fast and reliable convergence of the WLS estimation.

The procedure described before relates to a proper estimation of a power system state if all other required values are known *a priori*. However, there are series of other procedures that need to be performed before and after this estimation in the narrow sense, in order to provide the final information about the system state. Hence, when perceiving the problem from a broader point of view, it includes an on-line, automated tool in charge of providing a reliable, real-time database of the system, based on which all functions of Energy Managements Systems (EMS) are performed. Such state estimators typically include the following functions (not necessarily performed in the following order):

1. **Topology processor**
2. **Observability analysis**
3. **State estimation**
4. **Bad data processing**
5. **Parameter and structural error processing**

This thesis proposes new paradigms envisaged to cope particularly with 1^{st}, 3^{rd} and 4^{th} issue. Therefore, the following sections will provide an exhaustive introduction and a literature overview of these state estimation functions.

2.2 Gross Errors Processing - an Overview

In situations where measurement errors are independent random variables that are distributed according to the Gaussian distribution with zero mean and some degree of known variance, the WLS state estimator is undoubtedly the most appropriate estimator.

Unfortunately, it is possible for measurements to contain an error outside of this distribution which can substantially affect the final WLS estimation result. Therefore, the use of the WLS estimator is most appropriate under the assumption of Gaussianity. However, it does not possess any robust features that would allow identification of outlier measurements which may occur due to multiple reasons: meter failure, wrong meter connections, telecommunication system failure, interference, etc. Along with the outlying measurements acquired from power and voltage meters, anomalies can also occur in any data conveyed as an inputs to the state estimation procedure. Hence, all the errors that appear as outliers in the input dataset are in this thesis addressed as the ***Gross Errors*** (**GE**). Input variables that are corrupted with gross errors are called the ***Bad Data***. According to the type of data that GE may corrupt, they can be classified as:

- **Analog measurements GE**
- **Topology GE**
- **Network parameters GE**

The work described in this thesis is primarily focused on GE in measurements as well as topological structure of the power system. In any respect, if the input data are corrupted with gross errors, the WLS estimator will not be able to filter them out which results in false estimation of the system state, significantly different from its true values. All the analyses performed hereafter consider some type of gross errors, so for the sake of clarity, one explicitly defines a Gross Error as perceived in this thesis:

- **GE in analog measurements** (active/reactive power flows and injections and voltage magnitudes) – error in the range of 20 - 200 standard deviations of the Gaussian-distributed errors.
- **GE in topology** – erroneous circuit breaker status, i.e. reported status is closed, while in fact it is open.

Regarding the GE in analog measurements, Distribution of GE is assumed to be uniform, as shown in Figure A.1 in Appendix A. A general task of a bad data analysis within a robust state estimator is to detect the presence of GE, identify them and then correct or simply eliminate. Some

specific GE scenarios are of particular interest when testing robust abilities of a state estimator. An exhaustive overview about processing GE in analog measurements is provided in the subsequent section.

Gross errors in network topology are defined with inconsistency between true system topology, and information about the topology reported by the SCADA system. More specifically, GE in topology of a power system is defined by erroneous or missing breaker status. An exhaustive overview about topology processing issues is available in section 2.3.

Generally, in case of any input data corruption, one can expect significant deterioration of the estimated state vector if it is provided by the WLS estimator. This is due to the fact that the WLS estimator does not possess any abilities of filtering the bad data or elimination of their impact on the obtained solution. This problem was first recognized in the early beginnings of state estimation in power systems, and to date, a large diversity of research work has been published in order to equip the estimator with additional functions which would enable it to successfully detect bad data, identify and eliminate their distorting effect on the state estimation result. Generally, gross errors processing in most of robust state estimators is composed of the same functions, defined as the following:

- ▷ **Detection** of the GE presence in the SE input dataset.
- ▷ **Identification** of the input SE variables corrupted by GE.
- ▷ **Elimination** of the bad data from the input SE dataset.
- ▷ **Quantification** of GE in the identified bad data, or in other words **Correction** of the identified bad data.

The bad data processing functions are mostly processed in the respective order as they are presented. Hence, detection is mostly performed first, just after (or rarely before) the SE procedure. Only in cases when GE are detected, the identification process starts, aiming to determine which of the particular input variables contain GE. After bad data are identified, they should be eliminated from the input dataset, before restarting the estimation process. An alternative, more demanding and more desirable solution, is to correct the bad data instead of elimination. If all this processes perform successfully, the final SE solution should be free of GE deteriorating impact, and with a certain approximation equivalent to the true state vector.

The following sections will aim to represent and briefly explain most of the relevant efforts in building a state estimator robust to any kind of GE.

2.2.1 Static State Estimation – an Overview

Some of the bad data may appear as obvious, for example, negative voltage magnitudes and power measurements which severely deviate from Kirchoff laws or measurements which result in several orders of magnitude difference from the expected value. These types of measurements can be eliminated prior to state estimation by a simple brute data inspection. However, not all bad data can

be detected and identified by using this approach. Therefore, in order to build a sufficiently general and robust state estimator able to successfully detect and identify any type of bad input data, a substantial variety of methodologies relying on the local redundancy of the system information were proposed over the last 40 years. This section discusses only gross errors typical for the analog SCADA measurements.

Hence, a definition of the robust state estimator, unlike a normal one, is that it is expected to remain unbiased despite the existence of outliers in the input dataset. When analyzing the developed methodologies, it can be observed that robustness is commonly achieved at the expense of computational complexity required by additional features that provides it. The type of GE analysis depends on the used estimation method and can be performed after or before the state estimation process [4]. There are also state estimators that incorporate GE processing as a part of the estimation procedure.

The earliest methods were mostly conceptualized as post-estimation techniques where a WLS estimator provided the state vector and bad data was processed according to the estimated values. Logically, the most valuable indicators of bad data were measurement residuals (equation (2.4)).

The pioneer in bad data processing - Identification by elimination via the Largest Normalized Residual Test (LNRT) was established together with the first state estimation procedure by Schweppe *et al.* in [2]. Based on statistical properties of residuals, a predefined threshold is used for detection of a presence of gross errors. In case of positive detection tests, measurements with largest residual are eliminated and the WLS estimation is repeated. This process is then iteratively repeated until gross errors detection test becomes negative, i.e. there are no more detectable gross errors in the remaining set of measurements. It is important to mention that residual test method relies on a certain level of local redundancy. More specifically, if a measurement is not redundant then it is critical which means that elimination of this measurement remain the system unobservable. The residual of a critical measurement is always equal to 0, so if it is infected with a GE, there is no static state estimator that can reveal it. A more detailed analysis of measurement error detectability and residual spreading regions can be found in [6] and [7].

It is of key importance to perceive the following: residuals are all zero-mean normally distributed, but their variances can significantly vary between measurements. More specifically, a variance of a residual will depend on measurement type, location, relevant network parameters as well as on the precision of the associated meter device. Residual variances and covariances are expressed via the residual covariance matrix Ω calculable as follows [1]:

$$\Omega = R - H(H^T R^{-1} H)^{-1} H^T \tag{2.12}$$

where:

- R measurements covariance matrix,
- H Jacobian matrix.

Diagonal entries of matrix Ω express variances of each residual, while off-diagonal elements signify interaction among particular measurements. Therefore, in order to make the largest residual

test more objective, Schweppe *et al.* [1] proposed residuals normalization in the following manner:

$$r_i^N = \frac{|r_i|}{\sqrt{\Omega_{ii}}}, \quad \forall i \in \{1,\dots,m\} \tag{2.13}$$

This way, the residuals are assumed to be distributed in accordance with the Gaussian distribution $r_i \sim N(0,\Omega_{ii})$ and normalized residuals will have a standard distribution $r_i^N \sim N(0,1)$, for all measurements $i \in \{1,\dots,m\}$. For bad data detection and identification, the normalized residuals are shown as a superior criterion when compared to common residuals. Moreover, the LNRT is proven to be impeccable for identification of a single gross error. However, when multiple gross errors occur simultaneously, a masking effect can occur and cause a GE misidentification, due to interaction among the residuals. This means that correct measurements can be identified as erroneous and eliminated, whilst some bad data remained in the measurements set and thus deteriorate the final estimation result. In general, distribution and spread of the residuals that correspondent to erroneous measurements will directly dependent on the type, location, local redundancy and number of measurements infected by GE. Accordingly, one can distinguish the following gross errors scenarios:

- Single bad data
- Multiple bad data
  - Multiple non-interacting bad data
  - Multiple interacting non-conforming bad data
  - Multiple interacting and conforming bad data

As early mentioned, the problem of a single GE is easily resolved with the LNRT method. The identification of multiple gross errors however poses a serious challenge for the robust state estimation, which after almost 40 years of research, still has room for improvement. Multiple GE can still be easily or barely identified, depending on their location, magnitude and sign. In this regard, the non-interacting bad data can be defined as significantly distant in electrical sense, with no significant interaction between the residuals irrespective of their values. This type is the easiest to be identified since it allows a practical separation of the problem into multiple single bad data.

Alternatively, measurements, whose corresponding off-diagonal elements in the matrix Ω are high, will have a significant mutual impact on their residuals. When gross errors affect these measurements, they are denoted as interacting bad data which means that the identification may be fairly more demanding. When bad data are interacting, the information about the magnitude and the sign of the error is no longer insignificant.

The two interacting gross errors can appear as consistent between each other, i.e. they are conforming. In the opposite situation, they will be denoted as non-conforming. In particular, GE which are interacting and conforming are most demanding for identification and still represent an insuperable burden for most of the robust state estimation methods. The LNRT method is especially very sensitive to the conforming case when residuals of bad data may mask each other

at the cost of a correct measurements which are then eliminated instead. Nevertheless, despite this well-known weak spot, and many new methods proposed in the meanwhile, LNRT still remains the most frequently used concept for bad data processing.

Ever since the state estimation was established, it continued developing in order to build a general, automated and a robust process which would take into account all possible uncertainties. During this period of development, a particular type of estimator was initiated by Merill and Schweppe [8] who had an idea to process gross errors through a proper state estimation process and not with *a posteriori*, iterative analysis of residuals. They introduced the first M-estimator in power system state estimation via modification of the quadratic to the so-called ***Non-Quadratic Criterion*** (**NQC**). The idea of this methodology differs significantly from the previous LNRT method, in a way that the identification-elimination procedure of gross errors is implanted in the state process itself and not performed subsequently. The quadratic function in the WLS estimator results in a fact that large residuals have a much greater impact on the result than small residuals. In a general definition of the NQC, the quadratic function is intended to be modified to: quadratic-tangent (QT), quadratic-linear (QL), quadratic-square-root (QR) or quadratic-constant (QC) function in order to naturally screen and suppress the outliers in relation to the point where quadratic function changes to non-quadratic. This type of an estimator is able to automatically detect measurements with rapidly growing residuals and suppress their influence on the state estimate. The methodology is lately further enhanced in [9]–[11], especially regarding the definition and manipulation with the break-even point of quadratic function. Most recent work recalling of the NQC for GE elimination can be found in [12] and [13].

Another variant of the objective function, which can also be denoted as non-quadratic, is known as the ***Weighted Least Absolute Value*** (**WLAV**), proposed by Irving *et al.* in [14]. The idea is to substitute the quadratic function with a linear function in order to enable bad data rejection properties. The significance of the gained robust performances was subsequently proven and more elaborated in [15]. However, Falcão *et al.* in [16] warned against certain bad data combinations when WLAV estimator is particularly sensitive and fails in identification of gross errors.

Generally, the main advantage of the non-quadratic estimators is in their simplicity and relatively low computational requirements. The transformation of the quadratic objective function aims to embed in the method a mechanism for bad data detection/identification. Hence, the estimation and bad data analysis are carried out in the same procedure thus avoiding the iterative re-estimation until all gross errors are eliminated. Alternatively, adoption of non-quadratic functions may result in problems of local minima and convergence. In addition, there is still a possibility of bad data misidentification due to the fact that there is no one-to-one correspondence between the residuals and measurements errors.

An alternative approach, known as the ***Hypothesis Testing Identification*** (**HTI**), was developed in [17]. Unlike the previous methods, it uses individual criteria for each suspected measurement. The idea is to exploit the statistical properties of each error estimates for the suspected measurements and via individual identification testing decide whether the corresponding measurement is erroneous or not. This method shows relatively high robustness and allows for

identification of all gross errors in a single step. Since normalized residuals are used for detection of GE and for creation of a list of suspected measurements, this method inherits the same weaknesses that are characteristic for the LNRT method. In addition, computational heaviness for the required matrices may impose significant limitations.

All the early methods for bad data identification were exhaustively reviewed and comparatively analyzed in [18]. Despite various pros and cons, which characterize each method, there is a common property for all the methods. More specifically, since all of them rely on measurements residual, they are very harmful and fail in identification of multiple conforming gross errors. Moreover, bad residuals may be masked when GE affect measurements called ***leverage points***. It is recognized that some measurements, due to meter location and network parameters, create an outlier in the factor space, spanned by the rows of Jacobian matrix, and therefore have an undue influence on the estimation result. For example, GE will have vastly more serious impact on the result if appear in injection power measurements, then if power flow is infected. In addition, a leverage point may be created by a power line that has multiple times lower impedance comparing to other lines.

Leverage points that contain gross errors are referred to as ***bad leverage points***. In these cases, they are able to ruin any state estimator that minimizes the sum of certain function of residuals. In order to overcome these issues, substantial work has been done to identify and limit such deteriorating behavior of bad leverage points and thereby make classical estimators more robust. The background idea of achieving this, aims to modify the objective function in order to balance the influence of measurements irrespective of their location and type. This problematic was early discussed in [9]. Subsequently, it was applied in [19], [20] in order to solve the problem of high sensitivity of WLAV estimator to leverage points. The use of weights for bounding the influence of leverage points on a M-estimator results in the so-called *Generalized M-estimator*, or ***GM-estimator***. The first estimator of this type, eventually referred to as the *Schweppe-type GM-estimator*, was proposed in [9], with an intention to down-weight only bad leverage points. The robustness of this estimator was improved on the basis of projection statistics in [21], in the formulation of the so-called *Schweppe-Huber GM-estimator*.

Another group of state estimators was developed specifically to eliminate leverage point effects and enable identification of multiple conforming GE. These methods were mostly adopted from general robust statistics developed by Rousseeuw and Huber in [22], [23]. The first step in this sense was performed by Monticelli *et al.* in [24] who perceived bad data identification as a combinatorial optimization problem and adopted a branch-and-bound algorithm from the decision theory in order to solve it. A subsequent step came much later, from the similar perspective, but with the principle of minimizing the median of residuals. An example of this estimator type was the ***Least Median of Squares*** (**LMS**) estimator which was initially developed by Rousseeuw in [25]. The LMS estimator was then extended to nonlinear power system problem by Mili *et al.* in [26], providing a substantial improvement to state estimators of that time, which were prone to a masking effect of multiple bad leverage points. More specifically, they emphasized the effects of bad leverage points and proposed a new estimator, which yielded the correct solution even if half

the redundant measurements were bad leverage points. Given that this is the maximum amount of bad data that a robust estimator can process, the a new group of estimators was set by LMS, called the *high-breakdown state estimators*.

Another high-breakdown estimator representative was proposed in [27], as an estimator that minimizes the sum of the smallest ordered squared residuals up to the position k, the so-called ***Least Trimmed Squares*** (**LTS**) estimator. The LTS and LMS make the same family of high robust estimators, but they are also characterized by a quite high computational requirements. In addition, they both suffer from slow convergence rate, lack of scalability to large-scale systems and problems in situations where some collinearity exists in the measurements.

After LMS and LTS, there were many other attempts to build high break down methods for the same paradigms, by adopting more sophisticated methods in order to focus the exhaustive approach and improve robustness but never at the expense of computational performances. For example, Gastoni *et al.* [28] proposed a procedure which evaluated different state vectors obtained by solving observable critical sets from the available measurements. In order to speed up the selection among samples, a genetic algorithm is used and the optimal solution is considered to be the one with the maximum agreement between the remaining measurements and those in the sample. In [29] a similar approach was broadened in the sense of generalized state estimation and solved by a heuristic tabu-search meta-strategy. In [30], the same combinatorial approach is solved by particle swarm optimization. One more high-breakdown estimator worth mentioning solves inequality constraints constructed to model the uncertainty in the measurements - the ***Maximum Constraints Satisfaction*** (**MCS**) estimator [31]. Furthermore, a very similar approach called ***Least Measurements Rejected*** (**LMR**) was also proposed by Irving in [32], [33]. The LMR estimator minimizes the number of measurements that remain beyond the pre-defined tolerance ranges for each measurement. Unlike the MCS estimator, where the same formulation is solved by a genetic algorithm, the LMR estimator uses mathematical programming.

More recent work, based on robust statistics can be found in [34]. Unlike other methods, this method is used as a pre-processor by performing a random sampling and minimization of the LTS criterion. After some measurements are discarded, the final estimate of a system is found as WLS solution of the remaining measurement set. In order to solve a combinatorial problem, the authors have proposed a heuristically-determined number of samples which needed to be analyzed.

It can be observed that across all of these estimators, highly robust features are always gained by engaging a more exhaustive estimation process which requires not so modest computational costs. Despite many new methodologies, which are being adopted in order to facilitate such an exhaustive estimation of high-breakdown methods, there is still no uniquely accepted solution.

Table 2.1 provides a comparative summary regarding most representative properties of some of the most known robust state estimators. Generally, robustness of the methods is quantified by the concept of breakdown point, proposed in [35]. A breakdown point of an estimator is defined by the maximum number of measurements that can be affected by bad data that still do not deteriorate the estimated state vector. It is calculated by increasing the number m_b of arbitrary large gross errors until the maximum norm of the caused bias in the estimated state

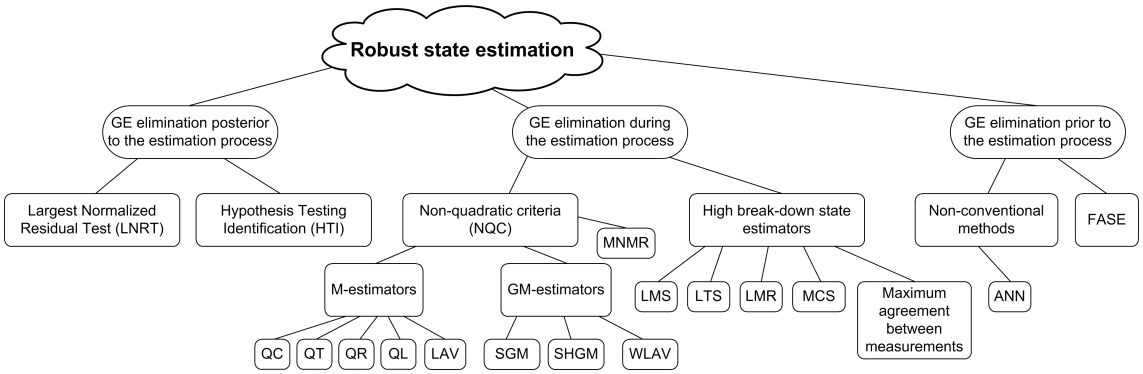


Figure 2.1: A principle classification scheme for most of the known state estimators. The classification is performed in accordance with the type of the used methodologies and the time necessary for bad data processing.

$b^{m_b} = \max_i |\hat{x}_i - \hat{x}'_i|$, $i = \{1, \ldots, n\}$ remains bounded. Therefore, $\hat{x}_i$ is a component of the estimated state vector before the introduction of GE, and $\hat{x}'_i$ is the estimate after the measurements vector z is corrupted with GE. Finally, the largest ratio $\frac{m_b}{m}$, (m is a number of available measurements) is finally called the breakdown point of an estimator. For example, in case of the WLS estimator, already a single GE will not leave the bias of the state vector bounded. Therefore, the breakdown point of the WLS is equal to 0. The highest level of robustness can be measured for the high-breakdown estimators which may remain immune when number of bad data is equal to a half of the number of redundant measurements m^r. This is the maximum amount of bad data a robust estimator can theoretically cope with. If the amount of bad data would exceed the maximum amount, it would no longer be possible to distinguish the good from the bad data.

In addition, estimators are compared in Table 2.1 according to some key properties: robustness level, problem definition type and pros/cons. Specifically the character of the mathematical problem, how most of the estimators are defined, can be a classical non-linear problem (NLP) or a mixed-integer non-linear problem (MINLP). It mostly depends if a binary variable vector is added to the optimization problem or not. From characteristics of the methods from Table 2.1, it can be observed that benefits of one perspective mostly include the limitations of the others. Most typically, robustness is gained at the expense of additional computational requirements.

In summary, Figure 2.1 gives a schematic, simplified classification for most of the mentioned state estimators. The three classes are defined according to the time of GE processing (detection, identification and elimination/correction) regarding the proper estimation process: posterior to SE, during the SE process or before the SE process (pre-filters). The estimators are named using abbreviations defined in this section. Hereafter, some alternative, unconventional and more recent robust state estimators are addressed.

As previously mentioned, most of the proposed classical robust state estimators are prone to failure due to the masking effect of bad leverage points. This particular problem was recently addressed by Bretas *et al.* in [38]–[41]. In order to keep the quantity of new information contained in measurements, the authors proposed indexes of undetectability and innovation based

Table 2.1: Comparative table with characteristics of the most relevant existing state estimators.

Estimator	Continuous /binary unknowns	Problem type	Robustness (breakdown point)	Pros	Cons	References
LNRT	$n/0$	NLP	$\frac{1}{m}$	- Simple - Easy to implement	- Prone to masking effect with multiple interacting GE - Computationally heavy due to iterative re-estimation/detection	[2]
(W)LAV	$(n+m)$ / 0	NLP	0	- Fast (no detection and re-estimation) - More robust than the WLS	- Very sensitive to bad leverage points	[14], [15], [36]
QC	n/m	MINLP	$\frac{1}{m}$	- Very fast and computationally efficient due to rejection and identification that are performed in the same process - There is no need for detection and re-estimation	- Demanding shape of the objective function with multiple local minima - Free parameter of break-even point - Prone to failure with multiple GE	[9]–[13]
HTI	$n/0$	MINLP	$\frac{1}{m}$	- Treats in parallel multiple GE - Corrects bad data instead of elimination	- Computationally heavier then QC and WLAV - Prone to failure with multiple GE	[17], [37]
LMS	$(n+1)/m$	MINLP	$\frac{0.5m^r}{m}$	- High robustness - Immune to bad leverage points	- Computationally heavy	[25], [26]
LTS	$(n+m)/m$	MINLP	$\frac{0.5m^r}{m}$	- High robustness - Immune to bad leverage points	- Computationally heavy	[27], [34]
MCS/LMR	n/m	MINLP	$\frac{0.5m^r}{m}$	- High robustness - High accuracy - Immune to bad leverage points	- Computationally heavy	[31]–[33]

on topological and geometrical approaches. These new measures are not reflected in residuals and thus may reveal gross error which could be masked by the state estimation process. However, in order to make this method applicable to processing of multiple gross errors, a significant upgrade would still be required.

Instead of a combinatorial search through multiple gross errors scenarios, He *et al.* [42] proposed an approach of modeling uncertainty among measurements. According to objective function of the proposed *Maximum Normal Measurement Rate* (MNMR) estimator, it brings it closer to the family of non-quadratic criteria. If compared to the estimator which satisfies the maximum constraints (MCS), this estimator substantially reduces computational requirements and becomes applicable to large-scale systems. However, the proposed model is non-convex and exhibits problems in finding the global optimum, common to most of the non-quadratic criteria.

A series of recently published papers by Caro *et al.* [43]–[45], which intend to modify the well-known WLS to *Dependent Weighted Least Squares* (DWLS) are also important to be mentioned. The key thought is that consideration of independence among measurements errors is inappropriate, especially if measurements are acquired from the same substation. In [43] the authors suggest a way to take measurement dependencies into account. Here, active and reactive power measurements provided by the SCADA are actually computed from the primary raw measurements: voltage magnitudes, current magnitudes and current-voltage phase angles. In other words, voltages, currents and phase angles, which are directly measured, are affected by statistically independent errors whilst power injections and flows, which are fabricated out of them, are obviously affected by dependent errors. It is shown that a proper model of correlation among measurements errors (as proposed in [44]) is able to make significant evolution in estimation quality. Furthermore, in [45] the same paradigm is particularly analyzed for the gross errors scenarios, when only a single bad measurement of voltages or currents can result in a measurement vector that contains multiple bad data. Accordingly, the authors propose to modify the LNRT on the way to consider the statistical correlation between measurements and make a significant improvement in multiple bad data identification.

The first installation of Phasor Measurement Units has started a revolutionary period, in both theory and practice, of state estimation in power systems. Since the first installation of devices equipped to measure the electrical waves (current and voltage phasors) at rates of several dozen times per second synchronously, the accuracy of power system monitoring and control is expected to substantially evolve. In general, any type of an additional measurement in a system improves redundancy and consequently the quality of state estimation. The PMUs, thanks to their accuracy and synchronicity, are able to make the most significant impact. However, this type of measurements is still finding its place in power systems, mostly due to economical burdens. Therefore, it seems kind of a transitional phase is to begin, when a fusion between traditional SCADA measurements and the new PMU measurements will be required. Regarding the issue of bad data processing, it is not necessary to emphasize the importance and potential of synchrophasor measurements. This challenge for evolving the state estimation robustness was already accepted by Chen and Abur in [46], [47]. Subsequently, efforts for opening a new era

in bad data processing by including PMUs were continued in [48]–[51]. The synchrophasor measurements are also very implementable under the light of the dynamic state estimation concept as proposed in [52]–[54]. Therefore, intense research work can be expected for development in this direction, which will set a path towards a full acceptance of new measurement standards.

Another actual research front in power system state estimation area is aimed towards development of a state estimator for distribution grids. A key important issue is concerned with scalability of the methods regarding dimensionality of the problem and processing of excessive data sets. Moreover, there are also some specific issues regarding implementation within the smart-grid concept. In general, the architecture of a state estimator in such an environment is expected to be decentralized and distributed to some local system entities. A significant recent contribution towards decentralization of the SE process, decrease in computational requirements and improvement in algorithm stability can be found in [55], [56]. A two-level SE concept is proposed, which at the first stage performs preprocessing of raw measurements at the substation level by a linear estimator, and then proceeds only manageable subset of measurements for the second level. A related work, which additionally considers availability of PMUs, is published in [57], [58]. Also a two-level state estimator is proposed, where decentralized local processing, on the level of a substation, is performed at the second stage, with the main scope of removing bad data and topology errors. Furthermore, paradigms for a multilevel state estimation, suitable with the smart grid environment are reported in [59]. Also, a substation state estimator for smart distribution grids is proposed from the same group of authors in [60]. Some more publications in the relevant area can be found in [61]–[63].

In this thesis, the state estimation problem is approached from an information theoretic point of view. Put simply, the input dataset is considered to be a general source of information, regardless of whether it was provided by SCADA or PMUs. For the purposes of simplicity, just SCADA measurements are discussed in the remainder of this thesis. Hence, all paradigms proposed in this thesis are designed to extract information from data, and therefore can be easily applicable to any measurement type.

After more than 40 years of state estimation experience in power systems, continuous research work in this area still remains current. More specifically, the search for a unique solution of a state estimator which is highly robust, accurate, computationally efficient (fast), flexible and scalable to large scale systems is still very much active.

2.2.2 Dynamic (Forecasting-Aided) State Estimation – an Overview

All the state estimation approaches mentioned earlier can be identified as static-natured state estimators. This means that historical state estimates are not considered when treating a current measurement snapshot. In order to enhance the robustness of static estimators, some alternative estimation approaches were proposed, with intention to extract valuable information from a succession of static-states evolving over time. These types of estimators are known as *dynamic state estimation* (DSE), or lately more preferably called the ***Forecasting-Aided State Estimation***

(**FASE**). This approach of observing time evolution of the static-state by considering multiple measurement scans is of particular interest for bad data processing.

The idea of dynamic approach was recognized in the early beginnings of state estimation, with a motivation to improve the capabilities of bad data processing. The early implementations were mostly based on quite naive models, without real forecasting capabilities [64], [65]. Following simplified attempts to track static-state, time changes were incorporated into the process, defining the so-called tracking estimators [36], [66]. A significant step towards the use of dynamic approach for identification of anomalies in the state estimation was made when the idea of innovation processes, generated by differences between forecasted and measured quantities [67], was proposed. Further enhancements in using forecasting capabilities for detection/identification of grossly erroneous measurements, based on computation of innovations, are established in [68].

A substantial progression in forecasting-aided state estimation is considered to happen after the introduction of *Artificial Neural Networks* (ANN) and pattern analysis. The pioneers of configuring a real-time network via pattern analysis were Alves da Silva and Quintana [69], [70]. In [71]–[73], normalized innovations are proved to be more appropriate for the ANN input, when being used for data debugging purposes.

In [74], ANN were also proposed for bus load prediction, as a step of dynamic state estimator. Here, forecasting was perceived as a three-phase procedure: bus load forecast, bus power injection forecast and online load flow. Similarly, hierarchical estimation techniques were incorporated in [75]. More recently, a fuzzy control method was proposed as a competitive study with extended Kalman filters [76]–[78]. Interesting work, coping with the challenging tasks of gross error identification in critical measurements via FASE, is described in more detail in [79].

An extensive review of all the work that improves state estimation by forecasting capabilities regarding data redundancy, innovation analysis, observability and bad data processing, can be found in [80]. The second of this two-paper series [81] presents the issues of implementation of FASE in a real *Energy Management System* (EMS).

The FASE state estimation intends to extract valuable information from previous measurement snapshots. This can be perceived as an auxiliary service that undoubtedly improves the estimation quality and reliability in situations where input data are distorted or missing. However, a question about unpredictable, sudden changes in power system that are not conveniently addressed by none of the mentioned work, needs to be raised. For example, in case of major topology change which establishes a break with the past, forecasting methods can barely accommodate this scenario correctly. Thus, in a case when state estimation result is of key importance for an appropriate control decision, the forecasting process, instead of providing aid, may only induce additional errors.

2.3 Topology Processing - an Overview

Section 2.2 provides an overview of detection and identification of gross errors in analogue measurements of power and voltage acquired from the SCADA. Another type of input data,

essential for SE and provided by the SCADA, are digital data that report statuses of circuit breakers, i.e. define topology of a system. This information about network topology delivered to Energy Management System (EMS) from the SCADA is also prone to gross errors. Thus, an erroneous or misidentified breaker status may results in inappropriate or even dangerous control decisions. Therefore, the information about topology needs to be considered as any other input data, i.e. prone to errors and inconsistencies so need to be taken under consideration by a specific estimation procedure. From the earliest establishment of state estimation in power systems, relevant research work brought about a diapason of topology estimation methods that are able to detect, identify and correct erroneous breaker statuses. These methods can mostly be incorporated in any of the state estimators analyzed in the former section.
All topology processing methods can be basically classified according to the moment of execution in the following manner:

- ***A priori*** – filtering of all the input data prior to execution of SE procedure.
- ***A posteriori*** – topological errors are detected and processed subsequently to the SE process, mostly by analyzing the measurements residuals.

In order to detect/identify bad data in analogue measurements, associated residuals provided by the classical state estimation (bus-branch) model are statistically analyzed. For the digital topological data there is no such correspondence, because the bus-branch model does not provide means for an explicit representation of switches and circuit breakers. Consequently, most of the earliest ***Topology State Estimators*** (**TSE**) were intended to reconstruct topology errors according to their impact on residuals of analogue measurements. One of the first attempts to detect topological errors was proposed in [82], with a principal idea of exhaustive repetitive state estimation for all the combinations of statuses (ON or OFF) for the suspected branches. Most of the methods, subsequently proposed during the 80's [83]–[85], were also based on such heuristic procedures and on identifiability of impacts that topology misconfiguration have on measurements residuals.

All these methods can be classified as *a posteriori* (or post-processing), since they require previous state estimation execution in order to obtain measurement residuals. All of these methods have a common drawback. More specifically, the WLS state estimation is known to have converging difficulties in ill-posed cases, so it is not surprising that presence of topological errors seriously deteriorate, or even disable the estimation efficacy. Hence, if a state estimator fails to converge, there is no prerequisite for the topology analysis. Moreover, as discussed in previous sections of this chapter, normalized residuals are prone to misidentification of bad measurements due to masking effect that multiple incident gross errors may cause. The same effect may also mask the topological errors. In order to overcome such failures, the paper published in [86] suggested a step towards a dynamic model by tracking the topology changes over time. With the assumption that radical topology changes do not occur frequently, consideration of the previous bus/branch model configuration may significantly save some computational time.

A contrary approach to all existing methods of that time was proposed in [87] and was based on an idea to process system topology prior to the state estimation procedure. This approach marked the establishment of the first *a priori* method. The work presented a systematic procedure, intended to locally validate switch indicators and analog measurements within a substation. Subsequently, aware of aforementioned weaknesses of the a-posteriori methods, Singh and Glavitsch [88] proposed a stand-alone expert system, i.e. a rule-based algorithm that intends to emulate a power system engineer. Significant originality of this work is in taking advantage of temporal consistency of measurements in order to define a switch status. This technique was even implemented in the existing EMS in a central European control center, and the gained experience was reported in [89].

A substantial stride in enhancement of TSE was introduced by Monticelli through the idea of zero impedance branches [90], [91]. More specifically, branches of the proposed bus/branch network model do not only represent transmission lines and transformers but also circuit breakers. This model resulted in a paradigm of the so-called ***Generalized State Estimation*** (**GSE**) [92]. The generalized estimator includes full uncertainty model for system topology, network parameters and bad data processing. The methodology is conceptualized as a two-step estimation procedure. The first stage includes classical procedure, together with anomaly detection method. In cases when anomaly is detected, the suspected zone is localized and passed to the second stage. The second stage is activated just in cases when error detection is flagged, including detailed description of the suspected zone on the level of substation bus section. This means that all the circuit breakers statuses are explicitly considered through associated binary variables which are submitted to state estimation procedure. Thus, the existence of a breaker is modeled with a null voltage difference at its extremities or with a null power flow through it. This allows direct estimation of a breaker status through the common SE method. The two-level estimation procedure establishes two ways of topology processing:

- ▷ **Bus-branch model** - the relevant information for topology error analysis is obtained from the conventional bus-branch model,

- ▷ **Bus section-switch model or Split bus model** - detailed network model, including all the individual circuit breakers within a substation. This model is exclusively applied as localized on a smaller part of the network that is suspected to contain topology errors.

After the GSE was proposed, a series of subsequent papers inspired by the same paradigm were published. Abur *et al.* [93] implemented the WLAV state estimation on both levels of topology processing in order to enhance the robustness to gross errors. A parallel work from another group of authors who intended to develop the WLAV topology estimator is addressed in [94].

Any of the aforementioned post-estimation techniques suffer from same weakness coming from the fact that the second stage process is based on values estimated either through the conventional WLS [91], or through the LAV estimator [93]. Being aware of these issues, Mili *et al.* [95] proposed an a-priori robust Huber M-estimator for topology error identification. The Huber M-estimator is applied to decoupled real and reactive power models with detailed

substation representation, where state variables assume active and reactive power flows at one end of every branch.

In [96], Clements *et al.* reported their work on a state estimator with flexible system topology or even system splitting. The idea was to define switching branch statuses as equality constraints and associate them with Lagrange multipliers, formulating the second stage of estimation process as an optimization problem. Additional a-priori information is also included in order to avoid the problems regarding critical sets of status information and system splitting during topology error identification. Computational efficiency of this approach was lately significantly improved in [97]. More specifically, statistical characterization of Lagrange multipliers allows the use of hypothesis testing in order to identify topology errors so that a computational burden can appear in large numbers of hypothesis alternatives. The work presented in [97] therefore proposed hypothesis testing based on Bayesian statistics in order to avoid multiple runs of state estimator for hypothesis evaluation. More recently, the same research group showed in [98] that the entire topology error identification procedure can be devised entirely in terms of collinearity tests between Lagrange multiplier vector and columns of the corresponding covariance matrix thus spurring the use of more elaborate hypothesis testing procedures.

Another relevant approach, also inspired by generalized state estimation and estimation of topology variables, can be found in [99], [100]. In this model topological, binary variables are introduced directly in the model and are estimated simultaneously with analogue electric variables. The breaker status is represented by a continuous real variable in the range $[0,1]$. In order to force the solution towards integer values of 0 or 1, the equation $x(x-1)=0$ is imposed as an additional condition. Also, a fuzzy controller is used in order to adaptively tune the weights and allow fast convergence of the WLS model. Comparative study of the approaches reported in [96] and [99], [100] was prepared by common work of the two working groups in [101].

One of most recent works that aims to identify status of breakers prior to state estimation process, is addressed in [102]. The problem intended to be solved is defined as a well-behaved mixed-integer quadratic programming problem, which exclusively contains topology variables (binary variable that takes value 0 for open breaker or 1 for closed breaker). The method seems appropriate even in the case of substantial number of breakers observed on the substation topology level (split bus model).

All methods mentioned thus far, can be referred to as conventional. This is due to their relationship with classical state estimation and/or mathematical programming approach which requires a solution of the optimization problem.

A shift towards an unconventional point of view on the TSE was done by recognition of the potential of Artificial Neural Networks. A hint about the use of ANN in PSSE was already given in section 2.2.2 through a review of dynamic state estimators that usually consider errors in both analogue and digital electric variables. Da Silva *et al.* in [103], [104] and [69], [70] first introduced pattern analysis as a combined solution for the topological identification, observability analysis and bad data processing problems. The main idea is to train an ANN offline, in order to obtain mapping through which selected measurements and breakers statuses are associated within their

corresponding topologies. In the real time, the trained ANN receives actual data vector from the SCADA and it provides a result which indicates one of the considered topology variants. This way the method is able to provide a very fast and reliable use of verified data as inputs to a classical state estimator, without any previous interaction. Particular advantages of the proposed paradigm can be associated exactly with the weaknesses of conventional methods: there is no need for a system observability check; one is able to deal with multiple interacting bad data and able to correct critical measurements. However, there are difficulties in proper ANN training that need to be overcame in order to reach sufficient efficiency of the method.

Data debugging was further improved in [71] by proposing normalized innovations as inputs of ANN. Normalized innovations are proved to be more relevant as information for recovering data anomalies when compared to measurements and normalized residuals. This is even visually justified in [72]. The same group of authors improved the method in [73], in order to make it more suitable for the real-time environment and processing of gross errors in both topology and analogue measurements.

Kumar *et al.* [105] approached topology identification using ANN from a slightly different point of view. Through a comparative study of two ANN architectures the new model was proposed for topology processing and state estimation and was shown to work well even for non-Gaussian noise and in the presence of bad data. A few more papers addressing usage of ANN for topology error identification can be found in [106]–[108].

An unconventional approach based on fuzzy pattern matching is proposed in [109] for all the key processes of state estimation: topology identification, bad data processing and state estimation. Fuzzy c-means clustering and pattern matching was used, where fuzzy pattern vector is generated based on the analogue measurement vector. Topology identification and gross errors are detected from difference between the fuzzy pattern vector and the analogue measurement data.

The problem of processing power system topology and identification of a breaker status when it is erroneous or missing has historically been approached from various points of view. The solutions based on generalized state estimation [110], were found to be implementable in the real EMS. Nowadays, state estimation is turning also towards distribution systems. Moreover, it will have a key role in the running of new concepts of grid organization, known as *smart-grids* and *micro-grids*. In these environments, state estimation will have to be upgraded with some specific features in order to meet the new challenges.

2.4 Recent Trends in Power System State Estimation

More than four decades after Schweppe established state estimation for power systems [2], the same WLS estimator is still accepted as a standard in transmission systems. In regard to bad data processing, conventional methods based on measurements elimination according to largest residuals, are still not forsaken and are still being accepted for industry implemented solutions. Although substantial research work over the last 40 years on development of a robust state estimator, there is still no final solution with all of the following desirable properties:

- ▷ Minimal computational requirements,
- ▷ Robustness to multiple GE,
- ▷ Robustness to bad leverage points,
- ▷ Accuracy in state estimation,
- ▷ Detection, identification and correction of the topology errors,
- ▷ Flexibility to absence of input electrical signals.

Moreover, new trends in restructuring of power systems have also raised some new challenges, which together with advent of computing power, revived interests in upgrading state estimation performance. Some of the hot topics which attract most of researchers' interests are:

- ○ **Synchrophasor-assisted state estimation** – development of a hybrid SCADA-PMU state estimator that will exploit the benefits of still a few PMUs in order to improve precision and robustness of the state estimation process,
- ○ **Robust Generalized state estimation** – there is still no satisfactory solutions for the increase in demand for robustness, flexibility and computational efficiency of a state estimator,
- ○ **Distribution system state estimation** (**DSE**) – expanding state estimation to distribution networks implies the following issues:
  - load prediction and pseudo-measurements generation in order to restore observability,
  - system topology identification,
  - integration of distributed generation creates new conditions in active distribution networks,
  - handling of enormous amount of measurements obtained from newly installed meters,
  - a distributed, decentralized and localized state estimation,
  - suiting the state estimation for the smart grids concept.

The electric power systems worldwide are mostly undergoing some crucial operational changes. Until recently, exclusively driven as vertically oriented, they are now submitted to open-access regulative, liberalized electricity market trading and increasing acceptance of non-utility players. Such changes will mostly affect distribution networks. Specifically, instead of passive, radial networks, they are becoming each time more active and dynamic in structure. This means that distribution systems do not just distribute electrical energy towards the consumers, but also enable reception of energy produced from distributed generation, which implies a more dynamic and flexible distribution system as a whole. All these intentions resulted in a new conceptualization of power systems that is popularly called smart-grids.

The electric distribution systems of today are still mostly static in nature with quite predictable loads and voltage profiles. Generally, there is still no significant penetration of active components such as distributed generation. Such networks do not increase requirements for frequent reconfigurations of feeders, protection or controller settings under normal conditions. These circumstances resulted in an approach of "fit and forget", where very little remote monitoring and control is provided to system operators. Some new roles required by Distribution Managements System (DMS) in an active distribution system are:

- increasing automation levels,
- optimization of network configuration,
- minimization of operational costs,
- acceptance of high penetration of distributed generation.

All these tasks are barely manageable without a satisfactory monitoring system which is still quite adequate in distribution systems of today, making a main barrier towards the realization of a DMS with advanced distribution automation. However, full monitoring, as performed on transmission systems, cannot be expected due to scarce economical justification. In any case, a distribution state estimator would be an inevitable function of this type of DMS. It will be expected to process all available measurements and load models in order to improve the system observability and provide the most probable state in the system. Improving the accuracy of estimated state vector is the main objective of both the input data and the estimation method itself.

The objective of this thesis is to establish new paradigms which can be adopted to any state estimation procedure. Although it is not envisaged to deal particularly with issues regarding the state estimation in distribution systems, all of the proposed concepts are designed to be suitable and easily adapted to any state estimation environment.

The power system state estimation was established on the least squares criterion in order to find the most probable state of the system. This criterion is undoubtedly the optimal approach under the assumption of Gaussianity of measurements errors. Thus, the WLS is still widely accepted in industry since it allows an accurate and computationally efficient estimation, very suitable for an online use. However, problems begin to appear when dealing with bad data since the WLS is not a robust state estimator and it cannot eliminate deteriorating effect on estimated state vector. Consequently, a large spectrum of different approaches for bad data detection and identification was developed since the beginnings of SE problem definition. However, the problem has not to date received a satisfactory solution for all the desired, aforementioned performances. In general, there is always a trade-off between robustness and computational effort, which can be barely reconciled.

When talking about gross errors, one mainly thinks about the bad data in analogue measurements. However, erroneous data can also be hidden in any input data that describe a system. A part of this thesis will be particularly focused on bad data in breaker statuses, i.e. network topology. A fundamental prerequisite for any supervisory and control action in a power system is the knowledge of system topology, defined by the statuses of circuit breakers. Thus,

taking a topology state as it is received from the SCADA, one is at risk of highly biased estimated state vector and inappropriate or even dangerous control decisions. Therefore, it is necessary to perform a breaker status identification method, able to extract information about the topology from the available measurements of power and voltage.

The principal idea of this thesis is to perceive the state estimation problem as a problem dealing with information. The associated challenges are principally the following:

- ▷ Robustness to gross errors in analogue measurements,
- ▷ Estimation of a power system topology,
- ▷ Observability of the values that are being estimated.

The main contribution of this thesis is to establish new paradigms in power system state estimation, built from the perspectives of *Information Theory* (IT) and *Information Theoretic Learning* (ITL). More specifically, the state estimation problem is perceived as a problem of maximizing information that can be extracted from the available measurements. In other words, the informational content of measurements errors should be minimized. Moreover, the task of topology estimation is perceived as an extraction of information about system configuration from the available analogue measurements of power and voltage. In all of these processes, new metrics are adapted to state estimation from Information Theoretic Learning and signal processing which are expected to bring improvement to robustness and accuracy in the presence of gross errors.

Chapter 3

Power System State Estimation Based on Correntropy

The state estimation problem in power systems has traditionally been solved by the Weighted Least Squares (WLS) criterion. By minimization of the sum of squared residuals, one in fact minimizes the variance of probability density function (PDF) of the measurement residuals. This would be the optimal method if and only when the error distribution was Gaussian. Unfortunately, this is not always the case due to outliers which appear as gross errors in measurements acquired from the SCADA. Such bad data are able to significantly deteriorate the WLS result and contaminate the estimative of healthy measurements. In any case, the WLS cannot provide the recognition and isolation of outliers and therefore cannot be called a robust state estimator. The importance of gross errors processing was recognized from the early establishment of state estimation theory which resulted in a wide diapason of robust state estimators. More details about all the relevant achievements can be found in Chapter 2. In regard to all properties a robust state estimator is desired to posses, it can be concluded that after more than 40 years of research work, the search for a fully satisfactory method is still active.

This chapter proposes a new paradigm for the power systems state estimation, inspired by the main idea of perceiving the state estimation problem from the information theoretic perspective. More precisely, the objective is to establish a criterion to maximize the information extracted from the available measurements or to minimize the informational content of error distribution. Put simply, if the error distribution has no information, then the PDF corresponds to a Dirac function and all errors are equal to zero. In this case, a unique solution exists for state variables which explain all measured values perfectly, i.e. total similarity is achieved. The key idea is the following: instead of minimizing the variance of error distribution, the intention is to resort to the concept of Entropy of the error distribution and to minimize the informational content in the process. Here, unlike the traditional WLS estimator which deals only with second moments, the intention is to extend to all moments of error distribution.

In order to shed a new light on the PSSE problem using this perspective, the descriptors are recognized in the framework of ***Information Theoretic Learning*** (**ITL**). The ITL has established

mechanisms for estimating descriptors from Information Theory (IT) directly from the data, with intention to substitute the conventional statistical descriptors used in nonlinear adaptive filters, machine learning and signal processing. According to the reported experience in signal processing, potential characteristics that ***Correntropy*** function could offer are recognized. More specifically, Correntropy is proposed as robust measures of similarity between the vector of available measurements and the electric values which are being estimated. Due to its unique properties of changing perspective of similarity with the distance, significant robust properties are expected to be availed if it is possible to implement Correntropy as a SE criterion.

The aim of this chapter is to establish a state estimator that is based on Correntropy, and test all of its properties using representative case studies. A principal proof of this concept was already published in [111], announcing very promising results in a sense of efficient and robust state estimator. This research is continued here in order to develop a novel state on an industrially-acceptable level.

3.1 Building Blocks - Information Theoretic Learning Paradigms

The milestone of this thesis are the paradigms borrowed from the framework known as *Information Theoretic Learning* (ITL), established in the late 90's by the Computational NeuroEngineering Laboratory (CNEL) at the University of Florida, led by Prof. José Carlos Príncipe. By intending to substitute the conventional statistical descriptors of variance and covariance in machine learning and signal processing, the ITL was developed as a framework of non-parametrically adapt systems based on information theoretic descriptors (entropy and divergence).

Specifically, in machine learning and signal processing, the procedure of learning from examples is traditionally based on criteria like correlation or mean square error (MSE). However, these do not entirely answer the intuitively asked question: *how to extract the maximum of information that is contained in data?* The importance of these issues was first recognized from a theoretical point of view by Shannon [112]. He established the concept of Information Theory (IT) in order to deal with the problem of optimally transmitting messages over noisy channels. This paradigm was quickly well-accepted by the scientific and engineering communities, resulting in dramatic impact on the communication systems design. Questions about the optimal code for data, or about the maximum amount of information which can be transferred through a particular channel are finally answered by introducing statistical descriptors called information ***Entropy*** and ***Mutual Information*** (**MI**). Put simply, the entropy quantifies the information content conveyed by a single random variable and mutual information considers two or multiple information sources thus quantifying the gain of information about one signal, by observing the other one.

The main motivation which induced paradigms developed within the ITL was to adopt the criterion and descriptors from information theory to adaptive signal processing and machine learning. This is due to entropy and divergence that are not directly implementable in learning-from-examples scenarios.

This thesis recognizes the benefits brought by these paradigms in signal processing and machine learning, and their potential for adoption in the PSSE. The main thought is to minimize the entropy of estimation errors at zero, with respect to the state variables. The following section brings a more detailed theoretic background, with an emphasis on the concepts of key importance for this thesis.

3.1.1 Entropy, Mutual Information and Correntropy

In order to weigh the amount of information, Shannon proposed the uncertainty measure called Entropy [112]. Let X be a random variable which takes values $x_1, x_2, \ldots, x_N$ with probabilities $p(X = x_k) = p_k$, where $p_k \geq 0$ and $\sum_{x_k \in X} p(x_k) = p_k = 1, \forall k = 1, \ldots, N$. Then, the entropy of X it is calculated as a sum across the set of uncertainties in each message weighted by the probability of each message:

$$H(X) = -\sum_{k=1}^{N} p(x_k) \log p(x_k) \tag{3.1}$$

with an assumption that for $p(x_k) = 0$, $p(x_k) \log p(x_k) = 0$. In other words, the entropy measures the average amount of information contained in a single observation of the random variable X and, in fact, provides a number of bits that are required to express this information. It is hard not to note the similarity of this entropy with the one defined in physics. However, informational entropy is a property of the probability mass function, whereas entropy in physical sense means a property of the physical state of a system.

Coming back to aspects of modeling some real phenomena, the difficulty in entropy estimation is the unavailability or uncertainty about the PDF of a process. This was the main motivation to develop the ITL. The ITL provides a way to directly estimate the entropy from the data samples, without imposing assumptions about the PDF [113]. For this reason, Shannon's formulation of entropy is not found as suitable and therefore more generalized measures of entropy were considered. A team from CNEL in [114], [115] adopted Renyi's quadratic entropy as the most appropriate for establishing the learning machine architecture which does not require a priori assumptions about data distributions, but is simply calculated directly from the data. A general way of expressing the entropy is:

$$H(X) = \varphi^{-1}\left(\sum_{k=1}^{N} p_k \varphi\left(I(p_k)\right)\right) \tag{3.2}$$

where $I(p_k) = -\log(p_k)$ is Hartley's information measure. If $\varphi(x) = x$ is selected, it becomes Shannon's entropy, while if $\varphi(x) = 2^{(1-\alpha)}x$ it is identified with Renyi's entropy of order α [116]:

$$H_{R\alpha}(X) = \frac{1}{1-\alpha} \log\left(\sum_{k=1}^{N} (p_k^{\alpha})\right), \quad \alpha > 0 \tag{3.3}$$

Particularly, for $\alpha = 2$, it corresponds to Renyi's quadratic Entropy:

$$H_{R2}(X) = -\log\left(\sum_{k=1}^{N} p_k^2\right) \tag{3.4}$$

Renyi's quadratic entropy is assessed as most suitable for achieving performances of ITL paradigms, due to previously stated reasons. In case of the continuous random variable X with PDF $p(x)$, quadratic Renyi's entropy becomes:

$$H_{R2}(X) = -\log\left(\int_{-\infty}^{+\infty} p(x)^2 dx\right) \tag{3.5}$$

Therefore, in order to achieve a non-parametric estimation of entropy, Renyi's quadratic entropy is combined with the ***Parzen window method*** for non-parametric estimation of PDF, as proposed in [114]. The Parzen window method, proposed in [117], estimates the PDF by centering a kernel function on each observation. Thus, it looks at a point as being locally described by a probability density Dirac function, which is replaced or approximated by a continuous set whose density is represented by the kernel. If assuming a Gaussian kernel $G(\mu, \sigma^2)$ (equivalent notation $G_\sigma(\mu)$ is used), with the mean μ and standard deviation σ, the estimation of the PDF for set of N individual samples $\{x_1, \ldots, x_n\}$ is obtained as summation of individual contributions:

$$\hat{p}_X(x) = \frac{1}{N}\sum_{i=1}^{N} G_\sigma(x - x_i) \tag{3.6}$$

When Shannon's definition of entropy is used along with Parzen PDF kernel estimator, an algorithm becomes insuperably complex. Now, one can get an answer about why Renyi's quadratic function is particularly chosen: when the PDF function is replaced in (3.5) with the Parzen's Gaussian kernel estimative (equation (3.6)), a convolution of Gaussian functions appears, which eliminates the need to calculate the integral. A replacement of the PDF with the kernel estimative emerges:

$$\hat{H}_{R2}(X) = -\log\left(\int_{-\infty}^{+\infty} \hat{p}(x)^2 dx\right) = -\log \hat{V}(x) \tag{3.7}$$

where $\hat{V}(X)$ is in the ITL vocabulary called the ***Quadratic Information Potential Estimator*** (or simply IP):

$$\hat{V}(X) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\int_{-\infty}^{+\infty} G_\sigma(x - x_j) \cdot G_\sigma(x - x_i)\, dx \tag{3.8}$$

The obtained integral of the product of two Gaussians is actually convolution of two Gaussians, i.e. it is equal to a Gaussian computed at the difference of the arguments, with a variance equal to

the sum of variances of the two original Gaussian functions:

$$\hat{V}(X) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\int_{-\infty}^{+\infty} G_\sigma(x-x_j)\cdot G_\sigma(x-x_i)\,dx = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} G_{\sigma\sqrt{2}}(x_j-x_i) \tag{3.9}$$

Therefore, it also gives an answer to the question: *why is only Gaussian function used for kernel in the Parzen window method?* Other kernel functions do not result in such a convenient evaluation of the integral because the Gaussian maintains the functional form under convolution [113].

To summarize, the ITL approach combines (3.4) with (3.6) in order to obtain an operational function representing the entropy of the error's PDF that can be directly implemented into training algorithm [113]. When calculating entropy, one just needs to evaluate Gaussian functions at distances between pairs of samples. The IP concept resulted in the proposal of the Minimum Error Entropy (MEE) [118] (using quadratic Renyi's Entropy) criterion as an alternative cost function for supervised adaptive system training. The MEE criterion represents minimization of the average information content of the error signal. Due to demonstrated superiority over the conventional ***Minimum Mean Square Error*** (**MMSE**) criterion, it was quickly accepted for a new neural network training criterion.

The concept of entropy deals with the characterization of a single source of information. In order to cope with multiple sources, the second important descriptor of statistical properties is proposed in information theory called ***Mutual Information*** (**MI**). Let $X = \{x_k\}_{k=1}^{N}$ and $Y = \{y_k\}_{k=1}^{N}$ be two discrete random variables. The total decrease in uncertainty in X by observing Y is known as the mutual information between X and Y, defined by Shannon as:

$$I(X,Y) = \sum_{j=1}^{N}\sum_{i=1}^{N} p(x_i,y_j)\log\frac{p(x_i \mid y_j)}{p(x_i)} = \sum_{j=1}^{N}\sum_{i=1}^{N} p(x_i,y_j)\log\frac{p(x_i,y_j)}{p(x_i)p(y_j)} \tag{3.10}$$

where $p(x,y)$ is joint PDF of X and Y, and $p(x \mid y)$ is a conditional probability density function. In other words, mutual information can be seen as a gain in information about X obtained by observing Y. Also, if X and Y are two independent variables, then a property of additivity holds:

$$H(X,Y) = H(X) + H(Y) \tag{3.11}$$

where $H(X,Y)$ is called joint entropy of X and Y. The explanation for this is the following: independence between random variables means that each variable contains no information about the other ones. When there is some form of dependency, then the following holds:

$$H(X,Y) = H(X) + H(Y) - I(X,Y) \tag{3.12}$$

Therefore, mutual information between X and Y quantifies dependency between them (intersection between $H(X)$ and $H(Y)$) and is always greater than 0. The equation 3.12 can be also alternatively

expressed as follows:

$$I(X,Y) = H(X) + H(Y) - H(X,Y) \tag{3.13}$$

$$I(X,Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X) \tag{3.14}$$

Once again, the ITL proposes a combination of Renyi's quadratic MI with Parzen window method in order to enable MI to be used as a criterion for training a feature extractor [119].

After the ITL benefits by error entropy criteria brought out already numerous of applications, a new step forward was made more recently in [120] by extending the fundamental definition of the correlation function of random processes to a *generalized correlation function*. They established a link between entropy cost functions and Huber's robust statistics through a new function called ***Correntropy***. It is able to include the formation of both the distribution and the time structure of a stochastic process. In [121], Correntropy is extended to a so-called general case of similarity measure between two arbitrary random variables. Given the two scalar random variables X and Y, Correntropy is defined by:

$$\upsilon_\sigma(X,Y) = E[G_\sigma(X-Y)] = \iint_{xy} G_\sigma(x-y)p(x,y)dxdy \tag{3.15}$$

where G_σ is the Gaussian kernel with a variance of σ^2. Since the joint PDF $p(x,y)$ is unknown and only a finite number of data $\{(x_i,y_i)\}_{i=1}^N$ is available, Correntropy can be estimated as:

$$\hat{\upsilon}_\sigma(X,Y) = \frac{1}{N}\sum_{i=1}^{N} G_\sigma(x_i-y_i) = \frac{1}{N}\sum_{i=1}^{N} G_\sigma(e_i) \tag{3.16}$$

This definition expresses Correntropy as a measure of similarity, that directly indicates the probability density of how similar the two random variables are in a specific window controlled by the kernel size σ. In other words, the kernel bandwidth acts as a zoom lens which controls the observation window in which similarity is assessed [122].

Among many properties of Correntropy, it is important to note that the Correntropy estimate $\hat{\upsilon}_{N,\sigma}(X,Y)$ is an integral of Parzen estimate of the joint PDF $\hat{p}_\sigma(X,Y)$ along the line $x = y$ [122]. This means that it estimates the probability of X being equal to Y. Hence, when one changes the system parameters in any process in order to maximize Correntropy, one is maximizing that probability. Furthermore, Correntropy is an extension of the correlation concept but it involves all even moments of the random variable $X-Y$. Therefore, it is a measure of non-linear dependency and not just linear, like the correlation.

Furthermore, Correntropy induces a distance function in the input space called the ***Correntropy Induced Metric*** (**CIM**) [121]. For two random vectors $X = (x_1,\ldots,x_N)$ and $Y = (y_1,\ldots,y_N)$ the CIM is defined as:

$$CIM(X,Y) = [\hat{\upsilon}_\sigma(0,0) - \hat{\upsilon}_\sigma(X,Y)]^{1/2} = [G_\sigma(0) - \hat{\upsilon}_\sigma(X,Y)]^{1/2} \tag{3.17}$$

The CIM is associated with similarity in the sense of information theory: the smaller the CIM distance is, the more similar the two distributions become. Moreover, it has been proved that CIM has all the properties of a metric:

- non-negativity: $CIM(X,Y) \geq 0$
- identity: $CIM(X,Y) = 0$ if and only if $X = Y$
- symmetry: $CIM(X,Y) = CIM(Y,X)$
- triangle inequality: $CIM(X,Z) \leq CIM(X,Y) + CIM(Y,Z)$

In addition, it has also been shown that CIM is invariant for translation of kernels such as the Gaussian kernel. Finally, the properties of key importance for the work presented in this thesis is the following:

- ▷ when two points are close, CIM behaves like a **L2 norm**,
- ▷ after exiting this zone, it transforms to the **L1 norm**,
- ▷ for two points far apart, it approaches the **L0 norm**, i.e. it becomes insensitive to distance.

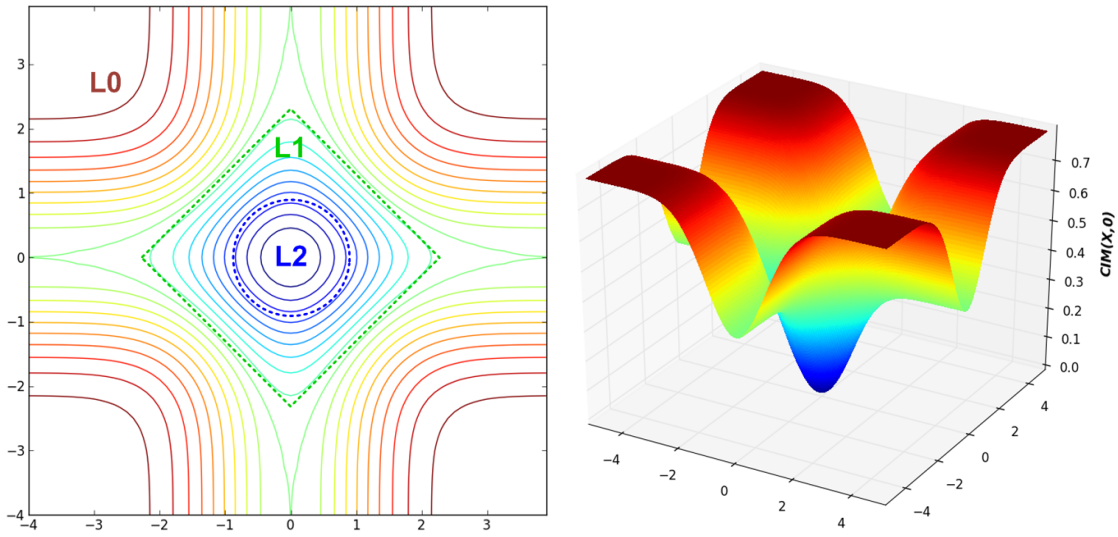
Figure 3.1, specifically depicts CIM distance of vector X from the origin in 2D and 3D projections. It is important to note the following: circular contours close to zero refer to the L2-norm; diamond contour is relevant to the L1-norm; the last contour after which CIM becomes constant, i.e. insensitive to the value of the error vector, corresponds to the L0-norm. Therefore, the scale of the CIM norm is directly controllable by the free parameter σ, i.e. size of the Gaussian kernel bandwidth. This way, the distance between two points is perceived relatively, in regard to the kernel size. Here it becomes obvious that the larger the kernel σ, the more similar CIM becomes to Euclidean distance. The proof of this statement is provided in Appendix B.1.

These appealing properties of Correntropy were suggested by [121] to be used as a criterion for adaptive systems training. Since the goal of the training process is to maximize similarity between the output and the desired signal, a new cost function is defined as the ***Maximum Correntropy Criterion*** (**MCC**). Let E be the error variable as $E = Y - X$ ($X = \{x_i\}$, $Y = \{y_i\}$ and $E = \{e_i\}, \quad i = \{1,\ldots,N\}$). The Parzen window method can be used to estimate the error PDF $p_E(e)$:

$$\hat{p}_E(e) = \frac{1}{N}\sum_{i=1}^{N} G_\sigma(e - e_i) \tag{3.18}$$

If PDF is evaluated at $e = 0$, and compared with (3.16) it states the following:

$$\hat{\upsilon}(X,Y) = \hat{p}_E(0) = \frac{1}{N}\sum_{i=1}^{N} G_\sigma(x_i - y_i) = \frac{1}{N}\sum_{i=1}^{N} G_\sigma(e_i) \tag{3.19}$$



(a) Contours of CIM(X,0) in 2D

(b) CIM(X,0) in 3D

Figure 3.1: (a): Contours of $CIM(X,0)$, i.e. distance to the origin projected on a 2D plane - notice how circular contours (Euclidean-like) change to squares (similar to L1) and to L0 (indifference). (b): 3D representation of $CIM(X,0)$ – notice the top regions of indifference in CIM value in distance from the origin ($\sigma = 1$).

Therefore, the maximization of Correntropy estimative has the effect of maximization of the value of the error PDF at zero:

$$MCC(E) \Leftrightarrow \max_{w} \sum_{i=1}^{N} G_{\sigma}(e_i) \tag{3.20}$$

where w denotes the parameters subjected to the optimization process and the control of the error.

The effect of optimizing the MCC criterion is to increase the value of the error PDF at the origin, aiming to approach a Dirac function in the ideal case. The most interesting fact about the MCC criterion comes from properties inherited from the CIM distance. More specifically, the MCC behaves like the minimum MSE criterion (or LS) when it comes to small errors. For the larger errors however it is transformed to the least absolute value (LAV) criterion in order to finally ignore the larger errors. Therefore, the perception about what is small and what is large is directly changed by the size of the Parzen window σ.

By recognition of robust properties that ITL concepts provide, this thesis intends to establish a new paradigm of information theoretic state estimation in power systems. Hence, the basic idea is to measure the information content of estimation errors PDF by its Entropy, instead of looking only at the variance of the error distribution. At the first glance, the most appropriate substitute of the least square error criterion seems to be the Error Entropy Criterion (EEC), i.e. the Minimum Error Entropy (MEE). However, when this criterion is linked in [113] with Huber's robust statistics, it resulted in the Error Correntropy Criterion (ECC), i.e. the MCC. Unlike MEE, the MCC does not

require centering of the errors mean. Since Correntropy is easier to calculate and since the goal of making the estimation errors equal to zero, the MCC is suggested as a more suitable criterion. As previously mentioned, the main motivation for building an information theoretic SE is in the fact that MCC shows major advantages when the errors PDF has a long tail, i.e. when it contains outliers in regard to the normal distribution. The following sections discuss the issues relevant to implementation of the MCC as a novel SE criterion.

3.1.2 Correntropy vs. MSE

For the purposes of better understanding of benefits the MCC brings in comparison to the conventional MMSE criterion, this section aims to clarify some of the basic properties, similarities and discrepancies between Correntropy and MSE.

The MSE for the two error variables $E = Y - X$ is defined as follows:

$$MSE(X,Y) = E[(X-Y)^2] = \iint_{x,y} (x-y)^2 f_{XY}(x,y)dxdy = \int_{e} e^2 f_E(e)de \tag{3.21}$$

where $f_{XY}(x,y)$ is joint PDF and $f_E(e)$ is marginal distribution of errors.

As can be seen from Figure 3.2(a) $MSE(X,Y)$ is a quadratic function in the joint space with a valley along the $x = y$ line. Therefore, the MSE represents a similarity measure since it quantifies the difference between X and Y. The distance in the whole space is L2-norm, i.e. Euclidean norm. Therefore, points which lay far away from $x = y$ will significantly contribute to the minimum MSE criterion and also impact the final optimization solution accordingly. This is the reason why the MSE performs well only in cases with normal distribution of errors and is not otherwise the an optimal choice.

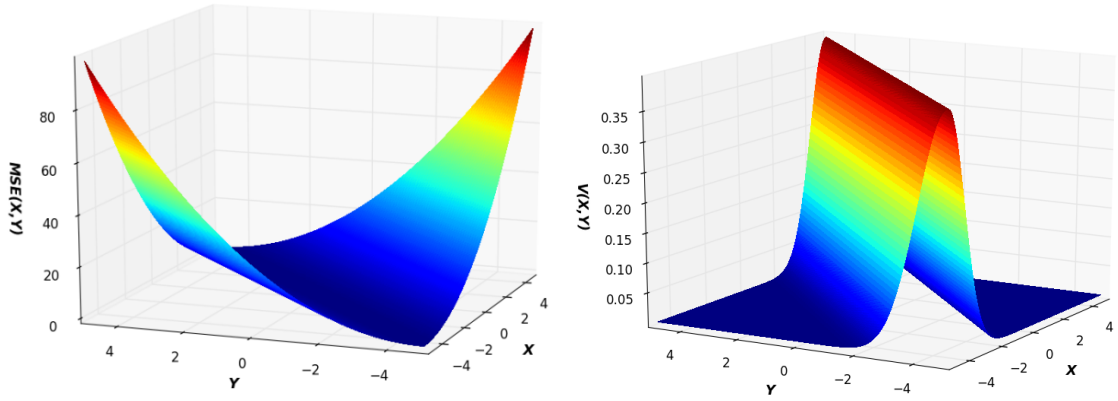
Lets recall the definition of Correntropy for the same variables X and Y:

$$\upsilon(X,Y) = E[G_\sigma(X-Y)] = \iint_{x,y} G_\sigma(x-y) f_{XY} dxdy = \int_{e} G_\sigma(e) f_E(e)de \tag{3.22}$$

As explained in the former subsection, the ITL proposes an estimation of Correntropy, directly from samples:

$$\hat{\upsilon}(X,Y) = \frac{1}{N}\sum_{i=1}^{N} G_\sigma(y_i - x_i) = \frac{1}{N}\sum_{i=1}^{N} G_\sigma(e_i) \tag{3.23}$$

As can bee seen from Figure 3.2(b), the Correntropy also represents a similarity measure in the joint space, but essentially in a different manner then the MSE. Along the line $x = y$ Correntropy has the highest value and can be correctly approximated with quadratic function. By moving away from $x = y$, the quadratic function changes to linear and the Correntropy value finally exponentially vanishes to zero. Exactly how far one needs to be in order to approach the zero Correntropy is directly determined by the size of the Parzen window, i.e. the bandwidth of the Gaussian kernel σ.



(a) MSE in the joint space X-Y

(b) Correntropy in the joint space X-Y with $\sigma = 1$

Figure 3.2: 3D plot of the mean square error (MSE) and the Correntropy function in the joint space $X - Y$.

In the case when size of the Gaussian kernel becomes sufficiently large, CIM distance is reduced to Euclidean distance, i.e. MCC criterion becomes equivalent to the minimum MSE criterion.

In regard to the comparison of how the MSE and Correntropy measure similarity, it is important to note one important issue: the MSE is a global measure, while Correntropy is local. Specifically, in the case of MSE, all of the samples will contribute to the objective function, where those which are far away from the expected manifold will contribute even more. On the other hand, in case of Correntropy, samples which do not pertain to the local zone will not take part in global maximum of the MCC cost function, i.e. they will simply be ignored. Locality of the observed area is controlled by the free parameter σ. In the case when Gaussian kernel is significantly enlarged, Correntropy will be equal to MSE in most of space. More details about the comparison between the MSE and Correntropy, as well as the proof of statement that correntropy with large σ corresponds to MSE, can be found in Appendix B.1.

According to this, it can be concluded that the Correntropy function possesses very attractive properties which can significantly enhance the training process of adaptive systems via the MCC criterion. In cases of poor Gaussianity, Correntropy will perform equivalently to MSE. On the other hand, when an outlier appears, the Correntropy has a natural mechanism to reject it while MSE remains biased by the large impact that outlier receives due to quadratic function. Accordingly, it can be said that with an appropriate control of kernel size, the MCC can never perform worse than the minimum MSE criterion. All these contributions of the novel MCC criterion were extensively corroborated with real samples of adaptive systems in [121] and [113].

These particular features that make the maximum Correntropy criterion robust to outliers, inspired the work presented in this thesis with an intention to incorporate the paradigms from ITL in power system state estimation. By approaching the state estimation problem from the information theoretic point of view, we will aim to adopt the MCC criterion. From experience

reported in theory of machine learning and signal processing, there is a reason to expect a respectable robust features from this type of MCC state estimator.

3.2 The MCC State Estimator – Implementation

This section proposes a novel information theoretic state estimator for power systems. The new paradigm is based on maximizing the information one can extract from the available measurements. Hence, the information content of PDF is measured by its Entropy and therefore, instead of only looking at the variance of error distribution, the new paradigm will look at the properties related to its entropy. The most appropriate background theory that gives answers to most of the questions raised by this informational view-point on PSSE are found in the information theory and the concept known as information theoretic learning [113].

The main idea is to define a SE criterion which would deal with the error entropy. A possible candidate in that sense could be the minimum error entropy (MEE) criterion. However, since it is independent of the distribution mean, it is definitely not suitable for implementation in the state estimation problem. The proper solution is found in the maximum Correntropy criterion. Hence, Correntropy will be used in order to measure similarity between measured vales and estimated measurements regarding the state vector. This section implements the new criterion and establishes the MCC state estimator. All properties of the novel state estimator are analyzed through preliminary case studies.

3.2.1 Adoption of the MCC Criterion in State Estimation Problem

Lets recall the state estimation problem, i.e. the overdetermined system of nonlinear equations:

$$z_i = h_i(x) + r_i, \quad i = \{1, \ldots, m\} \tag{3.24}$$

The aim of the estimation process is to find the most probable state vector $\hat{x}$ regarding the measured values in the vector z. The estimated measurements are calculated from the estimated state $\hat{x}$ vector as $\hat{z} = h(\hat{x})$. The errors from measured values are in fact estimative of the true metering errors, i.e. measurement residuals r.

The classical WLS state estimator searches for a state vector which corresponds to the minimal sum of squared residuals:

$$\begin{aligned} WLS(x) \Longrightarrow \min_x J_{WLS}(x) &= \min_x \sum_{i=1}^{m} \left(\frac{z_i - \hat{z}_i}{R_{ii}} \right)^2 = \min_x \sum_{i=1}^{m} \left(\frac{z_i - h_i(x)}{R_{ii}} \right)^2 \\ &= \min_x [z - h(x)]^T R^{-1} [z - h(x)] \end{aligned} \tag{3.25}$$

Therefore, the WLS in fact minimizes the normalized Euclidean distance between the estimated and the measured values. If one would consider dependencies between all the measurements, i.e. full errors covariance matrix R, such a dependent WLS would consider the Mahalanobis distance.

Since errors covariance matrix is assumed as diagonal, WLS is based on normalized Euclidean distance (or just Euclidean if all measurements variances have the same values).

This thesis proposes a new estimator based on perception of state estimation from informational point of view. In this sense, normalized Euclidean distance is substituted with the Correntropy Induced Measure (CIM) (eq. (3.17)) in order to measure the distance between the estimated and measured values. Accordingly, the estimator that minimizes the CIM distance is proposed, i.e.:

$$\begin{aligned} CIM(z,\hat{z}) &= (\upsilon(0,0)-\upsilon(z,\hat{z}))^{1/2} = (\upsilon(0,0)-E_{z\hat{z}}[G_\sigma(z-\hat{z})])^{1/2} \\ &= \left(G_\sigma(0)-\iint\limits_{z,\hat{z}} G_\sigma(z-\hat{z})p(z,\hat{z})dzd\hat{z}\right)^{1/2} \end{aligned} \tag{3.26}$$

where the second term $\upsilon(z,\hat{z})$ the Correntropy between measurements received from the SCADA z and estimated measurements $\hat{z}$. The ITL proposes the Parzen window method for the estimation of the PDF $p(z,\hat{z})$, so Correntropy could be evaluated directly from the samples:

$$\hat{\upsilon}(z,\hat{z}) = \frac{1}{m}\sum_{i=1}^{m} G_\sigma\left(z_i-\hat{z}_i\right) = \frac{1}{m}\sum_{i=1}^{m} G_\sigma\left(z_i-h_i(x)\right) \tag{3.27}$$

Since the first term in equation (3.26) $\upsilon(0,0)$ is a constant, the problem can be reduced from minimization of CIM to maximization of Correntropy. Therefore, instead of a maximum CIM estimator, a state estimator based on maximum Correntropy criterion is suggested:

$$MCC(x) \Longrightarrow \max_x J_{MCC}(x) = \max_x \frac{1}{m}\sum_{i=1}^{m} G_\sigma(z_i-\hat{z}_i) = \max_x \frac{1}{m}\sum_{i=1}^{m} G_\sigma(z_i-h_i(x)) \tag{3.28}$$

The novel state estimator is called the ***Maximum Correntropy Criterion*** (**MCC**) ***State Estimator***. It intends to maximize the similarity measure between the vector of available measurements and the vector of estimated measurements, i.e. maximize the errors PDF at zero. Hence, the new MCC objective function to be introduced in the PSSE theory for m measurements is the following:

$$J_{MCC}(x) = \frac{1}{m\sigma\sqrt{2\pi}}\sum_{i=1}^{m} e^{-\frac{(z_i-h_i(x))^2}{2\sigma_i^2}} \tag{3.29}$$

where σ_i is size of the Parzen window relevant for each measurement i (preliminary the same value $\sigma_i = \sigma$, $\forall i = \{1,\ldots,m\}$ is used). Since the amplitude of the kernel $1/(m\sigma_i\sqrt{2\pi})$ does not have any impact on the result of maximization process, the objective function $J_{MCC}(x)$ is reduced to the

following:

$$J_{MCC}(x) = \sum_{i=1}^{m} e^{-\frac{(z_i - h_i(x))^2}{2\sigma_i^2}} \tag{3.30}$$

According to all the properties that are to be inherited from Correntropy, the MCC objective function is expected to give new properties to PSSE that are particularly contributive to estimation robustness. A novel parameter of key importance is σ, a bandwidth of the Parzen window, i.e. a Gaussian kernel. When σ is substantially large, the MCC will behave equivalently to WLS (this is easily shown by taking the Taylor expansion of (3.30)). With smooth annealing of σ, the CIM distance will start to change from the L2 norm to L1 norm, i.e. the MCC may again be reduced to the LAV criterion. Further decreasing of σ finishes with the L0 norm. Therefore, if a particular measurement ends in the area of L0-norm it is in fact eliminated from the measurement vector, i.e. identified as bad data. Hence, the kernel bandwidth is a free parameter that directly impacts the shape of the MCC objective function and therefore regulates the impact of each measurement on the final estimation solution. If one looks at the equations (3.25) and (3.30), one could ask about the difference between the parameter σ which exists in both functions. The σ in WLS (eq. (3.25)), represents a standard deviation of measurements, their weights ($1/\sigma^2$). However, σ in MCC (eq. (3.30)) represents the size of the Gaussian kernel. Within the quadratic part of the MCC objective function, σ has the same meaning as in WLS, but it provides an additional role – it controls the break-even point where eventual outliers are to be eliminated. Such properties of the MCC estimator sound very promising and the following section intends to demonstrate this through representative samples.

3.2.2 The Proof of Concept

This section describes a few tests of the novel objective function for the purposes of examining its behavior and affirmation of the expected properties. Similar tests have already been presented in [111] where the concept of MCC state estimation is proposed and proved. The MCC criterion was initially attempted to be solved in [111] with a gradient ascent method which worked in the case of a toy problem (4-bus DC system). Larger system of 24 buses required a more powerful tool and the authors experimented with a meta-heuristic algorithm called the *Evolutionary Particle Swarm Optimization* - EPSO (more about this method can be found in [123]–[125]).

In this preliminary investigation of the MCC SE, the EPSO optimization will again be used to solve a simple state estimation problem while being aware that EPSO cannot be the final solution for the real-time state estimation. Here, the EPSO is used as a *brutal force* able to assure coming across the global optimum for most of the cases. The case study of a 3-bus DC system is chosen in order to visualize the objective function.

All the subsequent analyses will include results of normalized residuals r_N in order to be able to compare the MCC with the Largest Normalized Residual Test (LNRT). As already explained

in 3, the LNRT iteratively eliminates bad data according to the largest normalized residuals until bad data are no more detected.

In Figure 3.3, the 3-bus DC system (from [110]) is shown, with 3 measurements of active power injections: $P_1 = 4.06$, $P_2 = -4$, $P_3 = -0.02$ and 3 measurements of active power flows: $P_{1-2} = 2.02$, $P_{2-1} = -2$, $P_{1-3} = 1.96$ in $p.u.$ Variance of all the measurements is assumed to be $\sigma = 0.01 p.u.$

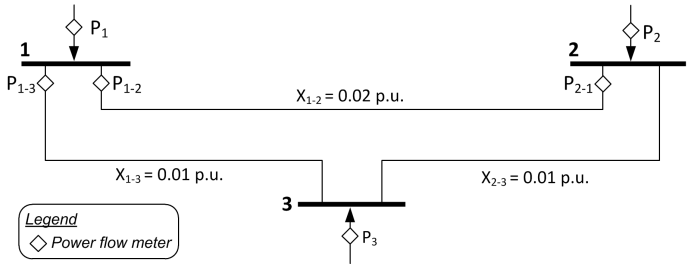


Figure 3.3: The 3-bus DC system.

In order to test the behavior of MCC state estimator in the presence of gross errors (GE), the following experiments were performed:

- Without GE,
- Single GE,
- Multiple interacting GE,
- Multiple interacting and conforming GE.

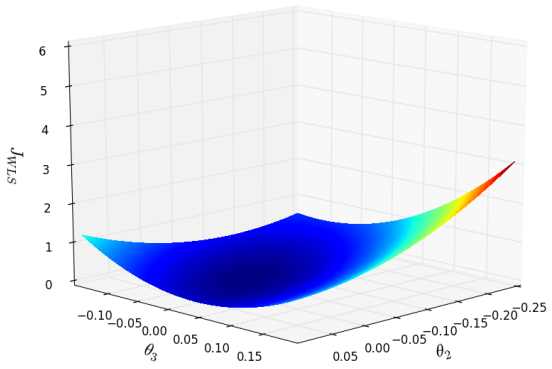


Figure 3.4: The objective function of the WLS criterion (J_{WLS}) for 3-bus DC system in the flat start ($\theta_2 = 0$, $\theta_3 = 0$) area. Convergence of WLS estimator is mostly fast and reliable, since the flat start is inside the convenient convex zone.

Without GE - this test sample assumes all the true measurements, i.e. all the measurements are free of bad data, but contain Gaussian noise with a presupposed standard deviation $\sigma = 0.01 p.u.$ which represents the accuracy of power meters. The results are obtained for the classical WLS and for the MCC while considering the sizes of the Parzen window $\sigma = \{10, 1, 0.1\}$. The lower bandwidth of a kernel is not reasonable due to assumed measurement standard deviation of 0.01. The MCC state estimation problem is solved by EPSO, while the WLS is solved by Newton's method (just a single iteration is enough due to a linearized DC problem).

The results from Table 3.1 show that differences between 3 estimated measurements vectors $\hat{z}^{WLS}$, $\hat{z}^{MCC}_{\sigma=10}$ and $\hat{z}^{MCC}_{\sigma=0.1}$ are negligible. A conclusion can be derived that measurement residuals are kept within the L2-norm of the MCC cost function. Consequently, the WLS and MCC for the both kernel sizes $\sigma = 10, 0.1$ are in this case practically equivalent and result in approximately the same estimated state vector $\hat{x} = [\theta_2, \theta_3] = [-0.0402, -0.0202]$.

The results are visualized by plotting the objective function in 3D, where the x-axis and y-axis mark the values of the voltage angles θ_2 and θ_3, respectively (node 1 is the reference node and therefore $\theta_1 = 0$ is assumed). Figure 3.4 shows objective function of the WLS criterion in the area around the flat start ($\theta_2 = 0$, $\theta_3 = 0$). Although one can only observe a single minimum, the WLS function in fact contains multiple local minima. However, the solution of WLS estimation is quite reliable in convergence to the proper solution, since the flat start is inside the convenient convex region, as it is obvious from Figure 3.4.

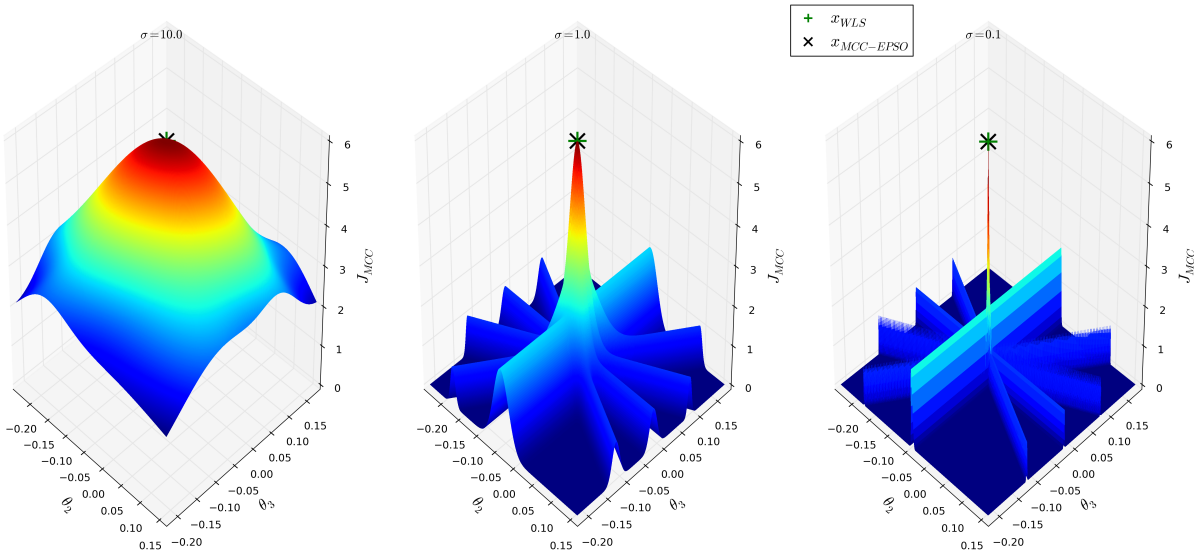


Figure 3.5: *Without GE* - The objective function for the MCC criterion (J_{MCC}) for 3-bus DC system considering $\sigma = \{10, 1, 0.1\}$. Notice the result of MCC remains the same and is equivalent to the result of the WLS despite the significant variation of the parameter σ. It has a single, global maximum that overlaps with the minimum of WLS.

Furthermore, it can bee seen in Figure 3.5 that the MCC global maximum overlaps with the WLS global minimum. For different Gaussian kernel sizes $\sigma = \{10, 1, 0.1\}$ the shape of the MCC

objective function significantly changes, however the maximum remains unique and unchanged. This confirms the fact that the MCC is equal to WLS until the residuals are sufficiently small in relation to size of Parzen windows, i.e. until they stay within the quadratic concave of the Correntropy function.

Table 3.1: *Without GE* - 3 bus DC problem – WLS vs. MCC. Results of state estimation with WLS and MCC with $\sigma = \{10, 0.1\}$. The difference between the 3 estimation results is negligible.

measur.	z^{true}	$\hat{z}^{WLS}$	$\hat{z}^{MCC}_{\sigma=10}$	$\hat{z}^{MCC}_{\sigma=0.1}$	r^{WLS}	$r^{MCC}_{\sigma=0.1}$	r^{WLS}_{N}
P_1	4.034	4.0273	4.0273	4.0273	0.0067	0.0067	0.91
P_2	-4.000	-4.0072	-4.0072	-4.0072	0.0072	0.0072	1.09
P_3	-0.010	-0.0201	-0.0201	-0.0201	0.0101	0.0101	1.56
P_{1-2}	2.011	2.0086	2.0086	2.0086	0.0024	0.0024	0.25
P_{2-1}	-2.001	-2.0086	-2.0086	-2.0086	0.0076	0.0076	0.80
P_{1-3}	2.025	2.0187	2.0187	2.0187	0.0063	0.0063	0.71

Single GE - This test sample assumes a single bad data in the injection measurement P_2. Results for the WLS and MCC with kernel sizes $\sigma = \{10, 0.1\}$ are given in Table 3.2. In this case, the MCC again gives the same result as the WLS for the large kernel bandwidth $\sigma = 10$. However, for smaller σ sizes the estimated measurements change and become more similar to true measurements. Put simply, it seems that unlike the WLS, the MCC with smaller kernel size simply remained immune to bad measurements. The largest normalized residual also correctly identifies the GE.

Figure 3.6 reveals how the MCC objective function changes by decreasing the kernel size $\sigma = \{10, 1, 0.3, 0.1\}$. It is significant to mention that for σ below cca. 1, the solution of the MCC x_{MCC} starts to differ from the WLS solution x_{WLS}. Finally for $\sigma = 0.1$, the solution provided by the EPSO is right on the global maximum of the J_{MCC} function. Unlike the case without the gross errors, it can be noted that maximum is no longer unique. However, some local maxima have emerged for $\sigma = 0.1$.

Figure 3.7 shows a bar of measurement residuals for the WLS and MCC with various Parzen window sizes. It is significant to note that with larger σ, a single GE results in a smearing effect of quite large errors in all measurements. On the other hand, for $\sigma = 0.1$, errors from all measurements are gathered around zero, despite the gross error in P_2. Also, with $\sigma = 0.1$, the size of error in the bad measurement is correctly estimated to cca. $-2.0 p.u.$ It can be concluded that by decreasing the kernel size, the result of the MCC becomes more similar to the true measurements vector z^{true}. Moreover, with sufficiently small kernel size, the MCC manages to naturally identify and practically ignore the erroneous measurement during the estimation process.

Table 3.2: *Single GE* - 3 bus DC system – WLS vs. MCC. Results of state estimation with WLS and MCC with $\sigma = \{10, 0.1\}$. It is important to note the distortion in the result, caused by a single GE in P_2. The MCC with $\sigma = 0.1$ managed to eliminate GE and provide a correct estimative of the measurement vector.

meas.	z^{true}	z^{bad}	$\hat{z}^{MCC}_{\sigma=10}$	$\hat{z}^{WLS}$	$\hat{z}^{MCC}_{\sigma=0.1}$	r^{WLS}	r^{WLS}_{N}	$r^{MCC}_{\sigma=0.1}$
P_1	4.034	4.034	4.4739	4.4747	4.0311	-0.4407	20.00	0.0030
P_2	-4.000	-6.000	-5.1377	-5.1388	-4.022	-0.8612	43.57	-1.9834
P_3	-0.010	-0.010	0.6638	0.6641	-0.0093	-0.6741	34.63	0.0044
P_{1-2}	2.011	2.011	2.4029	2.4034	2.0132	-0.3924	13.73	-0.0009
P_{2-1}	-2.001	-2.001	-2.4029	-2.4034	-2.0132	0.40237	14.08	0.0109
P_{1-3}	2.025	2.025	2.0710	2.0713	2.0179	-0.0463	1.74	0.0059

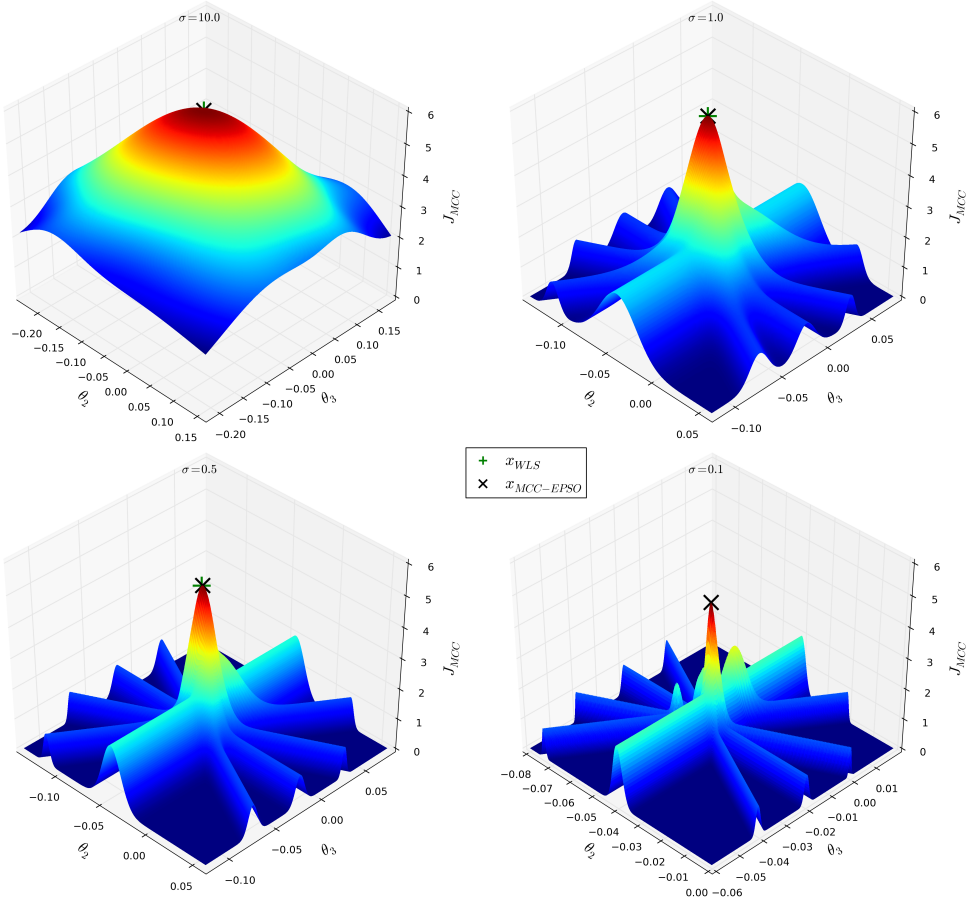


Figure 3.6: *Single GE* - an objective function of the MCC criterion (J_{MCC}) for 3-bus DC system considering $\sigma = \{10, 1, 0.5, 0.1\}$. It is significant to mention that for $\sigma = 10$, the global optimum of MCC is identical to the optimum of the WLS criterion. As σ decreases, the global maximum of MCC starts to differ from WLS minimum. Finally, for $\sigma = 0.1$ multiple local maxima can be noted. The EPSO however succeeded to find the global maximum, which corresponds to the solution which simply ignores the existence of bad measurement P_2.

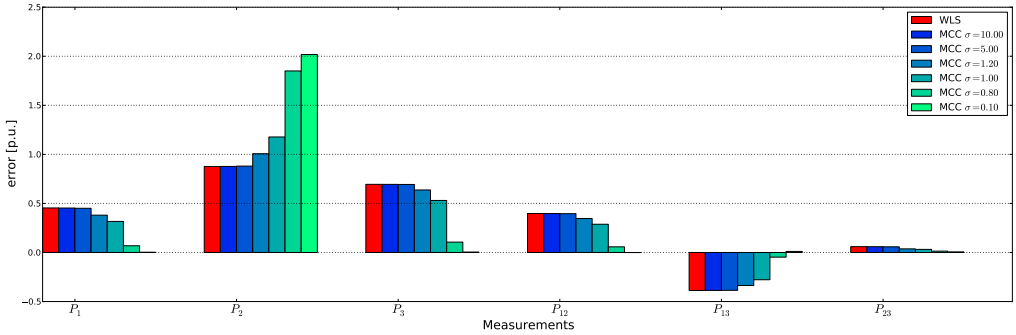


Figure 3.7: *Single GE* - A residual bar chart for WLS and MCC with various sizes of the Parzen window. Notice that by decreasing the Parzen window size, all residuals tend to zero except for bad data P_2.

Multiple interacting GE - in this test sample 2 gross errors are introduced in the measurements P_2 and P_{12}. Since these two measurements are incident and correspond to the same state variables, their residuals will be strongly correlated so they can attributed to the class of interacting gross errors. Detailed results in Table 3.3 show similar behavior of the MCC estimator in the case of single bad data. Obviously, it correctly identified both GE, similar to the LNRT according to normalized residuals.

Figure 3.8 clearly demonstrated that EPSO did not have any difficulty in finding the global maximum which corresponds to the solution which practically ignores both erroneous measurements, P_2 and P_{1-2}.

Table 3.3: *Multiple interacting bad data* - 3 bus DC system – WLS vs. MCC. LNRT and MCC manage to correctly identify both GE. The MCC with $\sigma = 0.1$ ignored both bad measurements and provided the correct result.

measur.	z^{true}	z^{bad}	$\hat{z}^{WLS}$	$\hat{z}^{MCC}_{\sigma=10}$	$\hat{z}^{MCC}_{\sigma=0.1}$	r^{WLS}	r^{WLS}_N	$r^{MCC}_{\sigma=0.1}$
P_1	4.034	4.034	4.1655	4.1655	4.0313	-0.1315	5.97	0.0027
P_2	-4.000	-5.000	-4.4743	-4.4742	-4.0171	-0.5257	26.59	-0.9829
P_3	-0.010	-0.010	0.3088	0.3087	-0.0142	-0.3188	16.38	0.0042
P_{1-2}	2.011	1.511	2.1599	2.1599	2.0121	-0.6489	22.70	-0.5011
P_{2-1}	-2.001	-2.001	-2.1599	-2.1599	-2.0121	0.1589	5.56	0.0111
P_{1-3}	2.025	2.025	2.0055	2.0056	2.0192	0.0195	0.73	0.0058

Multiple interacting and conforming GE - 2 gross errors are introduced in the measurements P_2 and P_{1-2}, but in case they support each other by their values, and are therefore called the conforming bad data. This case of interacting and conforming bad data is traditionally considered to be the most difficult bad data scenario to identify. It is can harm most of the robust state estimators, especially those that optimize a function of measurements residuals.

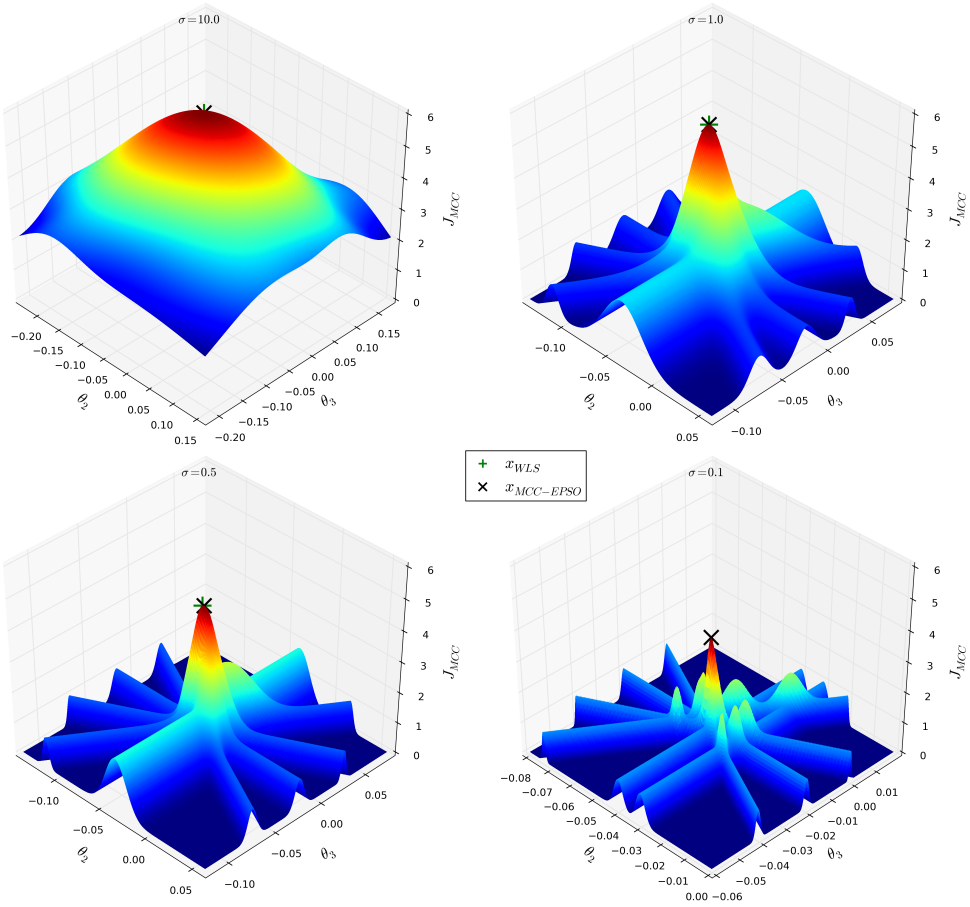


Figure 3.8: *Multiple interacting bad data* - objective function for the MCC criterion (J_{MCC}) for 3-bus DC system considering $\sigma = \{1, 0.5, 0.2, 0.1\}$. The global optimum of the MCC changes by decreasing σ from WLS solution to the solution that is free of GE impact.

The results for the WLS and MCC estimation are given in Table 3.4. Again, the MCC estimator using $\sigma = 0.1$ managed to successfully identify and reject bad data. On the other hand, the LNRT misidentifies bad data this way, i.e. it identifies correct measurements P_3 and P_{2-1} as erroneous due to the masking effect caused by the conforming gross errors.

Figure 3.9 shows a top view of the J_{MCC} objective function for $\sigma = \{10, 1.0, 0.3, 0.1\}$. The solutions of the EPSO optimization and the WLS solution x_{WLS} are also indicated. Again, one can note that for relatively large Parzen window 10 or 1 x_{WLS} and x_{MCC} are superposed. It is of key importance to observe the case with $\sigma = 0.1$, where x_{WLS} lays somewhere in between the local optima, while the state of the system provided by EPSO (x_{MCC}) is found to be on the top of the global optimum. The global optimum in this case corresponds to the solution which is obtained by WLS when both bad data are not taken into account. This result is very promising since it demonstrates that MCC estimator is able to be robust even in the most demanding GE scenarios and therefore it is expected to overwhelm most of the robust state estimators.

Table 3.4: *Multiple interacting and conforming GE* - 3 bus DC system – WLS vs. MCC. The LNRT fails to identify GE in P_2 and P_{1-2} thereby identifying the correct measurements P_3 and P_{2-1} as erroneous. The MCC estimator manages to correctly identify and reject both bad data.

measur.	z^{true}	z^{bad}	$\hat{z}^{WLS}$	$\hat{z}^{MCC}_{\sigma=10}$	$\hat{z}^{MCC}_{\sigma=0.1}$	r^{WLS}	r^{WLS}_{N}	$r^{MCC}_{\sigma=0.1}$
P_1	4.034	4.034	4.3365	4.3365	4.0313	-0.3025	13.73	0.0027
P_2	-4.000	-5.000	-4.6717	-4.6717	-4.0172	-0.3283	16.61	-0.9828
P_3	-0.010	-0.010	0.3352	0.3352	-0.0142	-0.3452	17.73	0.0042
P_{1-2}	2.011	2.511	2.2521	2.2521	2.0121	0.2589	9.06	0.4989
P_{2-1}	-2.001	-2.001	-2.2521	-2.2521	-2.0121	0.2511	8.78	0.0111
P_{1-3}	2.025	1.96	2.0845	2.0845	2.0192	-0.0595	2.23	0.0058

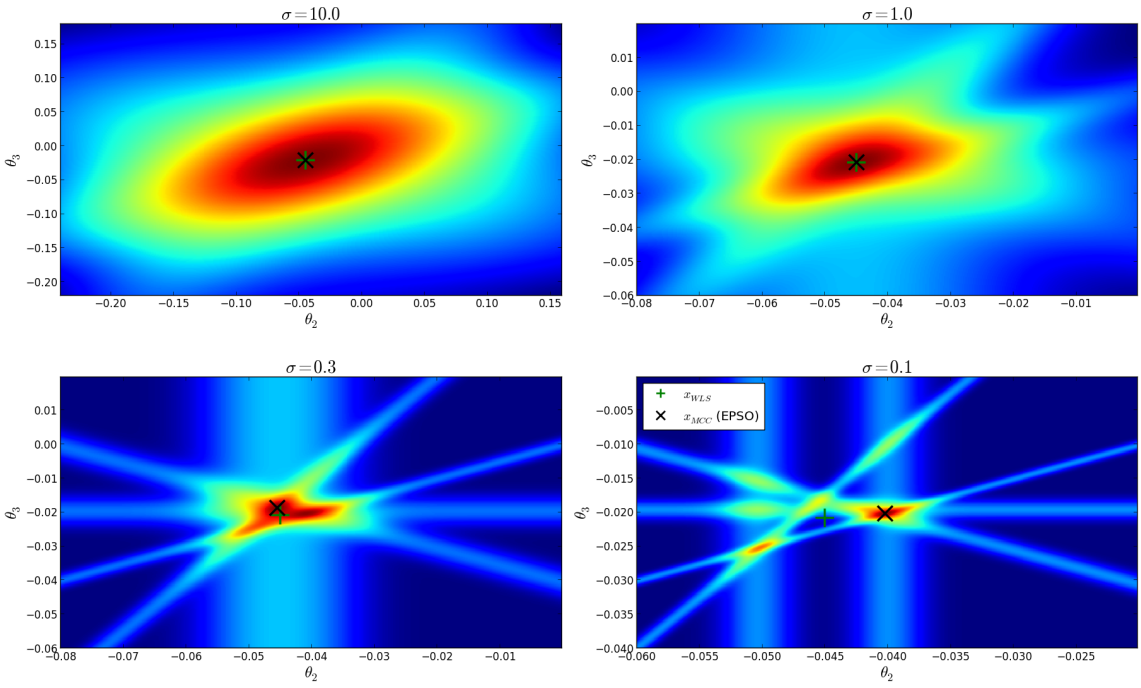


Figure 3.9: *Multiple interacting and conforming GE* - top view of the objective function J_{MCC} for 3-bus DC system considering $\sigma = \{10, 1.0, 0.3, 0.1\}$. For kernel size of 10 or 1 p.u., the global maximum of MCC is superposed with the minimum of the J_{WLS} objective function. For $\sigma = 0.1$, the EPSO manages to find global maximum and the WLS solution remains between the local optima. The global optimum of MCC corresponds to the solution that WLS estimator provides if both bad data are rejected.

In order to investigate the robust properties of the novel state estimation paradigm, more exhaustive tests on the same 3-bus DC system were performed. A 1000 scenarios were generated with all possible cases of 2 GE. More specifically, 2 measurements out of 6 are randomly chosen and turned to bad data by adding uniformly distributed gross errors in the range $[-0.5,-2.0]\cup[0.5,2.0] p.u.$ The MCC estimator with kernel size of $\sigma = 0.1$ and EPSO is

Table 3.5: The efficiency of LNRT and MCC-EPSO estimator in identification of 2 bad data in 1000 scenarios of the 3-bus DC system. The MCC method showed considerable superiority with impeccable bad data identification.

	Wrong	**Correct**	**Efficiency**
LNRT	140	860	86.0%
MCC-EPSO	0	1000	100.0%

compared with the method of identification by elimination in regard to LNRT. Each estimation case is evaluated as *wrong* or *correct* based on the success of GE identification. Table 3.5 reveals significant superiority of the MCC criterion over the LNRT method. As shown in relevant literature, the LNRT is prone to failure due to the masking effect of multiple conforming gross errors. On the other hand, the global optimum of the MCC criterion, achieved by the EPSO optimization method, corresponds to the solution which simply ignores the erroneous data. Accordingly, the MCC-EPSO estimator accomplished efficiency of 100% in identification of 2 GE in the 3-bus DC system.

3.2.3 The Solution Method

Although the remarkable results reported in the previous subsection were obtained by the MCC-EPSO state estimator, the EPSO cannot be seen as a permanent solution for solving the proposed state estimation based on Correntropy. Specifically, when larger systems with an AC estimation problem are analyzed, the EPSO is unfortunately too heavy in computational sense to perform an estimation feasible in real-time. Consequently, it cannot be considered as an industrially-acceptable solution.

In the proof of concept [111], a gradient ascent method is used (apart from EPSO) in order to maximize the MCC criterion for the 4-bus DC system case. Due to slow convergence and low reliability, this is also not an acceptable solver and a more suitable method needs to be found which would take into account the applicability in a real-time environment and also maintain the robust features of the MCC criterion preliminary announced in section 3.2.2.

In regard to these issues, it has been suggested to take advantage of the key property of Correntropy function that is differentiability everywhere and to every order. In that sense, the **Newton's method** has been proposed since it is fast, efficient and relatively easy implementable in maximization of the Correntropy criterion. One more argument which goes in favor of this choice is in the fact that the MCC criterion approaches the least squares, when the kernel bandwidth σ tends to infinity (Appendix B.1). In the remainder of this section, all issues related to adoption of Newton's algorithm in the MCC state estimation problem are discussed.

In order to maximize the Correntropy criterion, i.e. optimize the objective function $J_{MCC}(x)$

(eq. (3.30)), the first order optimality condition needs to be satisfied:

$$\nabla J_{MCC}(x) = \frac{\partial J_{MCC}(x)}{\partial x} = g_{MCC}(x) = 0 \tag{3.31}$$

$$g_{MCC}(x) = \frac{1}{\sigma^2}\sum_{i=1}^{m} e^{-\frac{(z_i - h_i(x))^2}{2\sigma^2}} (z_i - h_i(x)) \frac{\partial h_i(x)}{\partial x} = 0 \tag{3.32}$$

It can be written in the matrix form as:

$$g_{MCC}(x) = H^T(x) \cdot W(x) \cdot [z - h(x)] = 0 \tag{3.33}$$

where $H(x)$ is the Jacobian matrix from equation (2.6) and $W(x)$ is a $m \times m$ diagonal matrix with the i^{th} element defined as:

$$W_{ii}(x) = \frac{1}{\sigma^2} e^{-\frac{(z_i - h_i(x))^2}{2\sigma^2}} \tag{3.34}$$

If the function $g_{MCC}(x)$ is expanded into its Taylor series around the certain state vector x^k, the result is:

$$g_{MCC}(x) = g_{MCC}(x^k) + \nabla g_{MCC}(x^k)(x - x^k) + \ldots = 0 \tag{3.35}$$

When higher order terms are neglected from the eq. (3.35) it results in a iterative scheme of the Newton's method, i.e. correction of the state vector at each iteration k is obtained by solving the following:

$$\nabla g_{MCC}(x^k)\Delta x^{k+1} = -g_{MCC}(x^k) \tag{3.36}$$

$$x^{k+1} = x^k + \Delta x^{k+1} \tag{3.37}$$

The matrix $\nabla g_{MCC}(x) = \nabla^2 J_{MCC}(x)$ is known in PSSE vocabulary as the *gain matrix* $G_{MCC}(x)$:

$$\begin{aligned} G_{MCC}(x) = \nabla^2 J_{MCC}(x) = & \frac{1}{\sigma^2}\sum_{i=1}^{m}\left[\frac{1}{\sigma^2}(z_i - h_i(x))^2 - 1\right] e^{-\frac{(z_i - h_i(x))^2}{2\sigma^2}} \left(\frac{\partial h_i(x)}{\partial x}\right)^2 \\ & + \frac{1}{\sigma^2}\sum_{i=1}^{m}(z_i - h_i(x)) e^{-\frac{(z_i - h_i(x))^2}{2\sigma^2}} \frac{\partial^2 h_i(x)}{\partial x^2} \end{aligned} \tag{3.38}$$

As it was assumed for the classical WLS estimator in equations (2.10) and (2.11), the second term in eq. (3.38) can be neglected in order to avoid an additional evaluation of $mn \times n$ Hessian $\nabla^2 h(x)$. This simplification reduces the classical Newton-Raphson method to the ***Gauss-Newton***

method and the associated gain matrix takes the following matrix formulation:

$$G_{MCC}(x) = H^T(x) \cdot W(x) \cdot \left[diag \left\{ \frac{(z_i - h_i(x))^2}{\sigma^2} \right\} \right] \cdot H(x) \tag{3.39}$$

Finally, a corrective step Δx^{k+1} in the iteration form $x^{k+1} = x^k + \Delta x^{k+1}$ is calculated from:

$$G_{MCC}(x^k)\Delta x^{k+1} = -g_{MCC}(x^k) = -H^T(x^k) \cdot W(x^k) \cdot \left[z - h(x^k) \right] \tag{3.40}$$

$$\Delta x^{k+1} = -\left[G_{MCC}(x^k) \right]^{-1} \cdot H^T(x^k) \cdot W(x^k) \cdot \left[z - h(x^k) \right] \tag{3.41}$$

It should be kept in mind that the Newton's step is guaranteed to be in a descent/ascent direction if and only if $\nabla^2 J_{MCC}(x)$ is positive/negative definite. In other words, a search point of the Newton's method always needs to be kept within the convex/concave region in order to assure convergence to the local minimum/maximum. For the purpose of clarity, since the MCC objective function is to be maximized, a search point needs to be kept within the concave region in order to assure a negative definite gain matrix and ascent direction of the Newton's step. In case of the WLS objective function it is not necessary to be concerned with this issue, since the quadratic function is convex in the surrounding of the flat start (Figure 3.5), so Newton's method mostly converges to a unique solution in few iterations.

3.2.4 Adaptive Parzen Window Annealing

From the preliminary proof of concept in section 3.2.2, significant conclusions regarding the impact of the Gaussian kernel bandwidth on the shape of Correntropy function can be derived. In general, the size of the Parzen window directly controls the level of eventual outliers that are going to be ignored in the final solution. Accordingly, in order to eliminate all the GE, σ needs to be decreased to the level of expected Gaussian noise. In the cases free of GE, the solution of the MCC is expected to be equivalent to the minimum of the WLS cost function. Therefore, σ needs to be set to a value large enough to average the normal noise in measurements and small enough to avail the robust properties of the Correntropy function and eliminate the deteriorating effect of bad data.

In Section 3.2.2 the $J_{MCC}(x)$ objective function is set by the fixed, predetermined kernel size σ and optimized by EPSO. The following was deduced from preliminary MCC estimation and from visualization of the Correntropy function in the presence of GE:

- ▷ For the large σ, the Correntropy maximum is equivalent to the WLS minimum and no robust features are attained.

- ▷ When the kernel bandwidth σ is sufficiently below the gross errors level, multiple local maxima appear, but the global maximum corresponds to the true solution which ignores all GE.

Therefore, in order to establish a robust state estimator, the second case needs to be considered which is far more demanding due to multiple local optima. According to the reported results, the meta-heuristic optimization method - EPSO did not have any problems in finding the global optimum, even for the smallest acceptable σ and multiple local maxima. Unfortunately, the EPSO is assessed to be too heavy for real systems in the real time and Newton's method is barely feasible for such a demanding objective function with multiple local optima. Accordingly, one can conclude that predefinition of any fixed kernel size will always be a compromise between the optimization complexity and robustness to bad data.

As mentioned earlier, the most important issue one needs to keep in mind when optimizing using the Newton's method, is to assure a proper optimization direction, i.e. ascent or descent if one aims to maximize or minimize, respectively. The Correntropy function is maximized so the search point of Newton's optimization needs to always be kept within the concave region in order to assure a negative definite gain matrix and ascent direction of the step. Since Correntropy function is not negative definite on all of its domain, the ascent direction and quadratic convergence are not assured and caution is advised when applying the Newton's method to this function type.

The solution for this concern is found in the properties enabled by the free parameter, kernel size σ. Specifically, a large kernel size will over-smooth the surface of the function, i.e. it will make it quadratic and concave in most of the space thus allowing the the Gauss-Newton method to be easy applicable. However, the final intention is to end up with a minimal value of σ in order to attain the robust features of the MCC and reject the GE, if they exist. Therefore, the idea is to perform a smooth annealing of the Parzen window in parallel with the Newton's optimization. More specifically, the estimation procedure will start with a sufficiently large kernel size, and will then be smoothly and carefully annealed in order to avail the benefits of the MCC criterion to naturally ignore the bad data. Moreover, since the MCC with a large σ is proved to be practically equivalent to the WLS, one can also initiate the MCC estimation process with the WLS solution (instead of the flat start) and perform a continuous σ-annealing until it reaches the lowest acceptable value. If there were no GE, the MCC estimator should provide a solution equivalent to the one provided by the WLS. In the presence of GE, the WLS will be deteriorated, while the MCC is expected to remain immune to the same effect. At the first glance, the conclusion is that the MCC will never underperform the WLS.

The strategy of the Parzen window smooth annealing was also observed in the application of ITL paradigms in signal processing [126] where the need for a user independent annealing scheme has been emphasized. In a more recent report, Singh and Príncipe introduced an error-adaptive kernel which is large when errors are large and anneals as the error power reduces [127]. This way, the kernel width parameter is automatically updated online, achieving a relatively high gradient throughout the adaptation process and speeding up the convergence accordingly. For the sake of clarity, it is important to emphasize that the term *annealing* does not have any relation to *simulated annealing*. It could be observed that annealing is not the best name for the process of decreasing the size of the Parzen window. However, *Parzen window annealing* or *Kernel size annealing* is adopted directly from the Information Theoretic Learning, e.g. [126], [127]

The first tests of maximizing the Correntropy criterion with the Gauss-Newton method were performed by a quite primitive strategy of smooth kernel annealing where the kernel size was iteratively reduced in the annealing step n by a fixed value: $\sigma^{n+1} = \sigma^n - \Delta\sigma$. This strategy works in most of the cases, but it is not free of user adjusting, and may require significant computational resources.

Accordingly, a fast and self-adaptive annealing strategy is proposed which determines the kernel size in order to track the residuals of good (non-rejected) measurements. Therefore, a maximum residual of the non-rejected data is considered in order to calculate the new kernel size. Since the Gauss-Newton method needs to be kept within the concave area of $J_{MCC}(x)$, the border of Gaussian kernel concave area is taken as a reference point in order to calculate the relevant kernel size. Generally, the point where a concave function changes to convex or vice versa is called the saddle point, and in that point the second derivative is equal to zero. Therefore, the equality $G''_{\sigma_i}(z_i - h_i(x)) = 0$ is satisfied for the state vector x where residual absolute value is equal to the kernel size, $|r_i(x)| = |z_i - h_i(x)| = \sigma_i$. Therefore, the concave area of the Gaussian kernel is within the area $\langle -\sigma_i, \sigma_i \rangle$. Accordingly, an online update rule for the kernel bandwidth σ_i, for each measurement i, is proposed as:

if $|r_i(x^k)| \leq \sigma_i$
$\quad \sigma_i^{n+1} = \max_j |r_j(x^k)|, \quad j \in \{1, \ldots, m\} \setminus S$
else
$\quad \sigma_i^{n+1} = 0, \quad S \leftarrow i.$

where S is the group of suspected bad data. According to the proposed algorithm, measurements which are found outside the concave area of the associated Gaussian are identified as bad data. In order to maintain the concavity and assure the ascent step direction, a kernel size of the particular measurement needs to be abruptly reduced to a sufficiently small value, i.e. $\sigma_i^{n+1} = 0$. The proposed annealing scheme is completely autonomous and self-adapting. In the mean time, it improves the computational efficiency of the method and enables the preservation of the conditions required for application of the Gauss-Newton method. The kernel size is updated according to the proposed rule until it is larger then the smallest acceptable value σ_{min}. More specifically, if all the bad data are eliminated, the residuals will be reduced to the level of Gaussian noise that exist in measurements due to imperfection of the meters. Since all the measurements have these errors by definition, there is no reason to decrease the kernel size to that extent. Instead, it is desirable that the MCC behaves like the WLS since it is the best estimator in the presence of Gaussianity. Therefore, the σ_{min} is predefined to a value larger then the maximum expected value of residuals in cases free of gross errors and lower then the minimum expected gross error.

3.2.5 The MCC State Estimator - Algorithm

Due to complexity of $J_{MCC}(x)$ and proposed kernel annealing, the first experience of adoption of the Gauss-Newton method indicated some instability of the convergence rate, or even divergence.

The reason for this comes from well-known problems of the iterative step size, which in some cases may be too large and cause an undesired search progress. Hence, despite the proposed kernel annealing scheme, the search point may also jump out of the concave hull and even diverge. Thus, some adaptation of the iterative Newton's step is suggested, i.e. the so called line-search optimization, in order to perform an optimum search progress. Line search methods generally optimize the step size along the direction Δx^k, i.e. determine an optimum coefficient α to manipulate the step size:

$$x^{k+1} = x^k + \alpha \cdot \Delta x^{k+1} \tag{3.42}$$

Among various line search methods, it has been decided to implement quite a simple and computationally efficient solution called *Backtracking Line Search*. The termination criterion is defined with the so-called ***Armijo rule***, i.e. step size is being iteratively contracted until the following criterion is satisfied:

$$J_{MCC}(x_k + \alpha \Delta x^{k+1}) > J_{MCC}(x_k) + c_1 \alpha \left[\nabla J_{MCC}(x_k)\right] \Delta x^{k+1} \tag{3.43}$$

More details about line search algorithm and the Armijo rule can be found in Appendix C.1.

One of the main contributions of this thesis is the proposal of the Maximum Correntropy Criterion state estimator for power systems. In regard to the implementation of the MCC estimator in a power system, the following main components were discussed previously: the MCC criterion (Section 3.2.1 and 3.2.2), the Gauss-Newton method (Section 3.2.3), the strategy of Parzen window annealing (Section 3.2.4). Finally, a principle algorithm flow for the MCC state estimator is defined as follows:

The proposed MCC state estimation algorithm is conceptualized on alternating between the Gauss-Newton optimization and annealing of the kernel size. The Gauss-Newton is run until the maximum component of the iteration step vector Δx^{k+1} is lower then ε (typically $\varepsilon = 10^{-3}$). Each iteration step needs to satisfy the Armijo rule in order to refresh the current state vector: $x^{k+1} = x^k + \alpha^l \Delta x^{k+1}$. During the optimization procedure, checks are performed in order to determine if some of the residuals $r_i(x^{k+1})$ became larger then the associated kernel bandwidth σ_i^n. In that case, measurement i is identified as erroneous $S \leftarrow i$, and the associated kernel size is set to zero. After the Gauss-Newton stopping criterion is satisfied, the kernel size is annealed to the largest residual in regard the measurements which are not identified as erroneous: $\{1,...,m\} \backslash S$. Initial kernel size σ^0 is set to a value sufficiently large to make the MCC equivalent to the WLS estimator. The minimum kernel size σ_{min} is set to a value sufficiently large to perform as the WLS for the small Gaussian-distributed errors.

It is important to emphasize that the MCC SE, with the presented algorithm, will perform within the same procedure all the three important components of the bad data processing: detection, identification and elimination of GE, without requirements for any additional processing. Specifically, in the case when there are no GE, optimization with large σ^0, is expected to result in relatively low residuals. Accordingly, in the subsequent step, updated σ^1 will be lower

Algorithm 1 MCC state estimation

Define σ_{min} and ε
Initialization:
    iteration index $k = 0$
    initiate state vector x^0 to "flat start"
    set σ^0 to a large enough value in order to $J^{\sigma^0}_{MCC} \sim J_{WLS}$
while $\max(\sigma_i^n) > \sigma_{min}$ **do**
    while $\max|\Delta x^k| \leq \varepsilon$ **do**
        Calculate Δx^{k+1} (equation (3.40))
        $l = 0$; $\alpha^l = 1$
        while Armijo rule not satisfied (equation (3.43)) **do**
            $\alpha^{l+1} = 0.5\alpha^l$
            $l = l + 1$
        $x^{k+1} = x^k + \alpha^l \Delta x^{k+1}$
        if $|r_i(x^{k+1})| > \sigma_i^n, \forall i \in \{1, \ldots, m\}$ **then**
            z_i is identified as bad data $S \leftarrow i$
            $\sigma_i^n = 0$
        $k = k + 1$
    $\sigma_i^{n+1} = \max_j |r_j(x^k)|, \forall i, j \in \{1, ..., m\} \backslash S$
    $n = n + 1$
$\hat{x}_{MCC} = x^k$, the final MCC SE solution

than σ_{min} and the algorithm will terminate with a solution $\hat{x}_{MCC}$, equivalent to the solution provided by the WLS estimator $\hat{x}_{WLS}$. In this scenario when $\hat{x}_{MCC} = \hat{x}_{WLS}$, illustrated in Figure 3.5, it is guaranteed that input measurements are free of GE.

When GE appear, the estimation process will require a few more iterations to free the state vector from the GE impact, i.e. to identify and eliminate the GE. The MCC estimator in fact does not eliminate the bad data, it only annuls their impact on the estimation solution. This is essentially different from, for example, elimination of measurements according to largest residual test. In the LNRT, each measurement elimination requires recalculation of the relevant matrices and the observability test. Therefore, the proposed algorithm, which couples Newton's search and kernel annealing very efficiently incorporates all the functions required for bad data processing. The subsequent section includes case studies with exhaustive tests of the novel state estimator.

3.3 Results and a Preliminary Case Study

3.3.1 Case Study - the 3-bus DC System

The same 3-bus DC system case study, from section 3.2.2 , (Fig. 3.3) is used to test the proposed MCC state estimator. For the purposes of clarity, in section 3.2.2 , the Correntropy criterion with a fixed Parzen window is maximized by the meta-heuristic optimization method EPSO. Here, for the same Correntropy maximization, the annealing Parzen window is used in parallel with the Gauss-Newton method, as described through the presented algorithm.

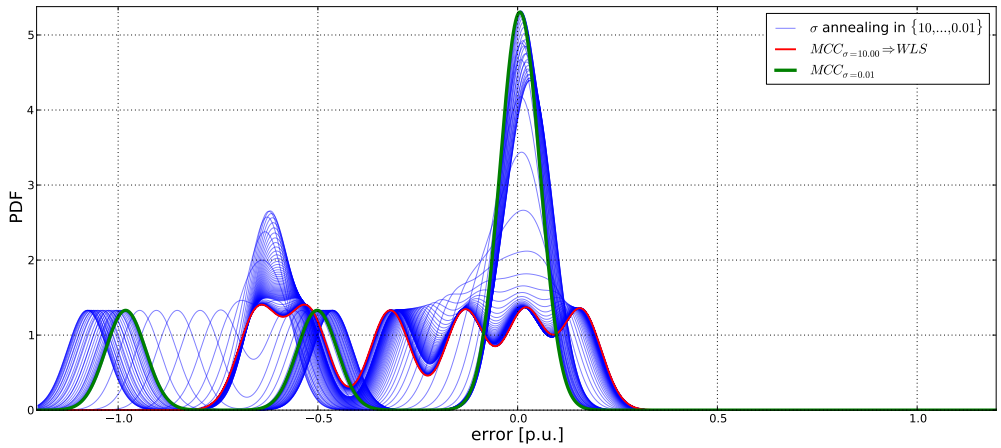


Figure 3.10: Evolution of the error PDF by annealing of the kernel size $\sigma \in \{10, \ldots, 0.01\}$.

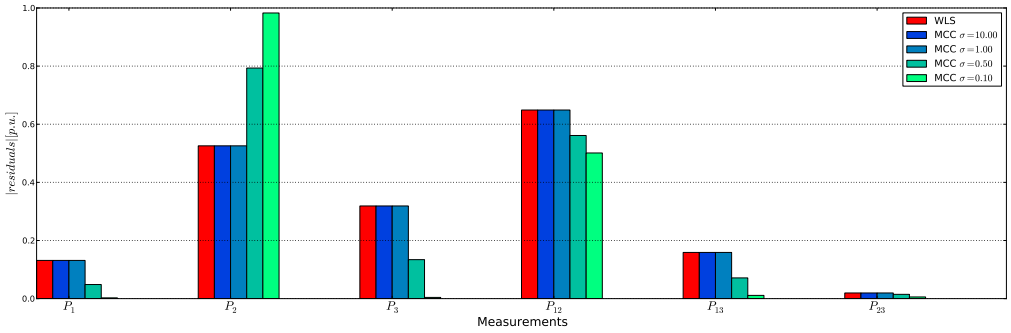


Figure 3.11: *Multiple interacting and non-conforming GE* - evolution of absolute residuals from WLS through MCC with $\sigma = \{10, 1, 0.5, 0.1\}$.

The typical scenario with 2 interacting and non-conforming GE is considered, as represented in Table 3.3. In order to be more illustrative, the Parzen widow is shrunk by a predetermined fixed step size ($\sigma^{n+1} = \sigma^n - \Delta\sigma$, $\sigma_i = \sigma$, $\forall i = \{1, \ldots, m\}$), and the residuals are calculated regarding true measurements. The evolution of residuals PDF for each Parzen window σ^n is shown in Figure 3.10. PDFs are estimated by the Gaussian kernel density estimation method using $\sigma_{PDF} = 0.05$. Starting with the large $\sigma^0 = 10$, MCC provides the same result as the WLS. When observing the associated PDF (red line), it is obvious that the GE had caused the so-called smearing effect among all the residuals. Afterwards, by a smooth decrease of the kernel size, residuals PDF is changed with an intention of condensating the true measurements around zero, while the 2 bad ones are being rejected (on the left-hand side). The final PDF (green line), obtained with $\sigma = 0.05$ indicates that all residuals are clustered around zero.

Therefore the 2 GE are identified and correctly estimated to values -1 and -0.5. The bar plot for the same residuals is shown in Figure 3.11. It is important to note that with large Parzen window, the residuals of the true measurements are also large, whilst the annealing kernel enables robust

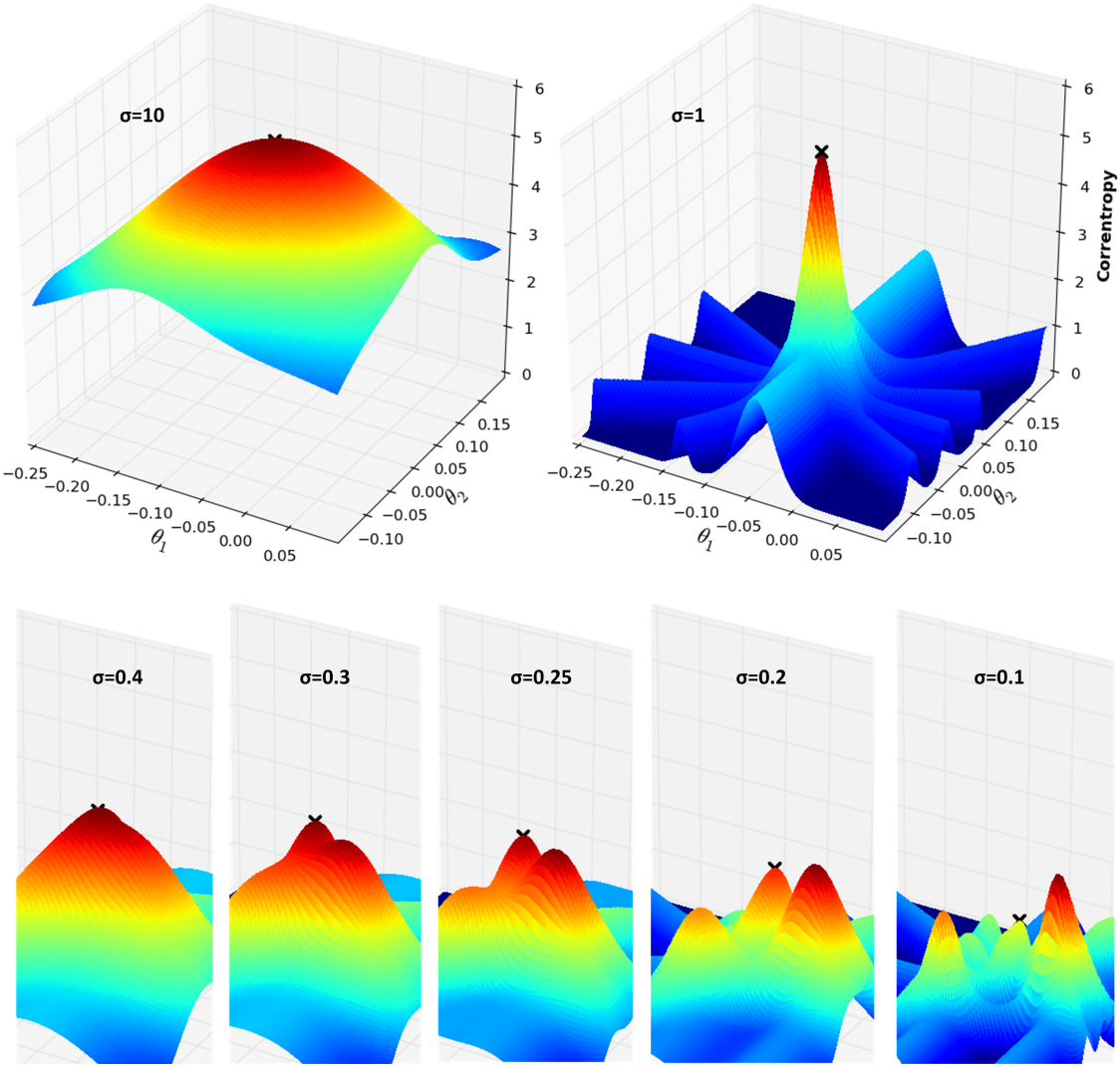


Figure 3.12: *Multiple interacting and conforming GE* - through the Parzen window annealing Newton's step is kept trapped in the concave region, which for the ultimate kernel size $\sigma = 0.1$, ends up as local maximum, rather than the global one.

features of MCC which identifies GE P_2 and P_{1-2}, and reduces residuals in all other measurements to almost zero. The 3D plot of the Correntropy function will be identical to the one from Figure 3.8 (the solution of the MCC estimation solved by Gauss-Newton is identical to the one obtained with EPSO).

The other typical scenario, observed in section 3.2.2, with 2 interacting and conforming gross errors (Table 3.4), is solved by the novel MCC state estimator. The 3D plot of the Correntropy function and the MCC SE solutions are shown in Figure 3.12 for 7 characteristic Parzen window sizes. Unfortunately, it can be observed that the iterative solution is trapped in the concave region, which for the ultimate kernel size $\sigma = 0.01$, ends up as local maximum, rather than the global one. In other words, the MCC estimator had misidentified the bad measurements and, at their cost, eliminated correct measurements, free of gross errors. While σ is shrinking, the objective function

landscape changes from a well behaved quadratic shape to a surface with numerous local optima. In this particular case, when σ is being reduced, a local optimum emerges in a location away from the concave region, where the current solution is located at that iteration step. Afterwards, a step of the Gauss-Newton method is not able to leave the concave hull and keeps the solution trapped in the continuation of the annealing process. Although the formed concave region of the new local optimum may rise to become the global optimum for narrower σ values, transition from the local maximum, where the Guass-Newton finds the final solution, is no longer possible.

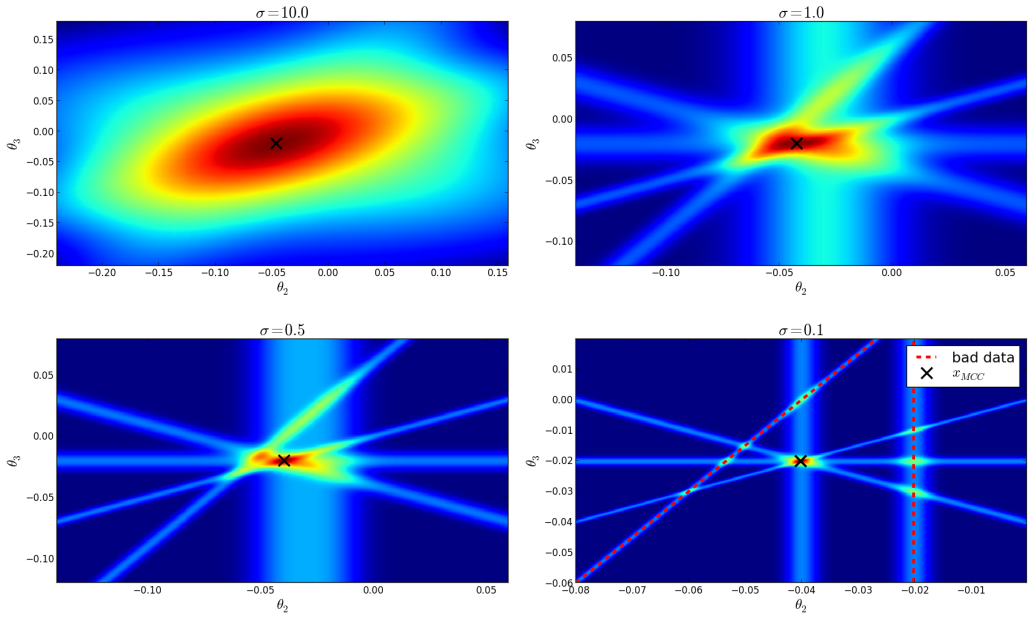


Figure 3.13: 2D projection plot of the Correntropy function with 2 interacting non-conforming GE $(-2,-1)$ in measurements P_2 and P_{2-4}. The MCC estimation is successful in finding the global maximum.

This specific failure sample is characteristic for the fact that the 2 GE considered are interacting by their location and conforming by their value. For instance, in the previous sample, the 2 GE were only interacting but not conforming and the Gauss-Newton did not have any problems in finding the global optimum. In order to observe the difference between these two cases, the 2D plots of the Correntropy functions for non-conforming and conforming GE scenario are shown in Figures 3.13 and 3.14, respectively. It can be clearly observed that the conforming GE $(-2,1)$ support each-other and in the earlier phase, when the Parzen window is still large, manage to attract the solution far from the true state vector. When the Parzen window is decreased to the level of emergence of local maxima, the solution remains in the local maximum and the MCC fails to identify and reject the GE. This effect is not present for the errors of values $(-2,-1)$ since they do not support each-other, i.e. they are not conforming and the MCC succeeds in identification.

It is important to observe that the same problem with interacting and conforming GE is well-known difficulty for the LNRT method. Specifically, the conforming GE cause the masking effect among the residuals, i.e. their residuals are decreased at the cost of residuals corresponding

to true measurements. This way, an ordered elimination of measurements with largest residuals results in GE misidentification. The same problem obviously impacts the Newton's search through the process of annealing kernel size of the Correntropy function. With large Parzen window, the MCC is equivalent to the WLS, so the Newton-annealing optimization is evidently mislead by the same masking effect to the undesired solution.

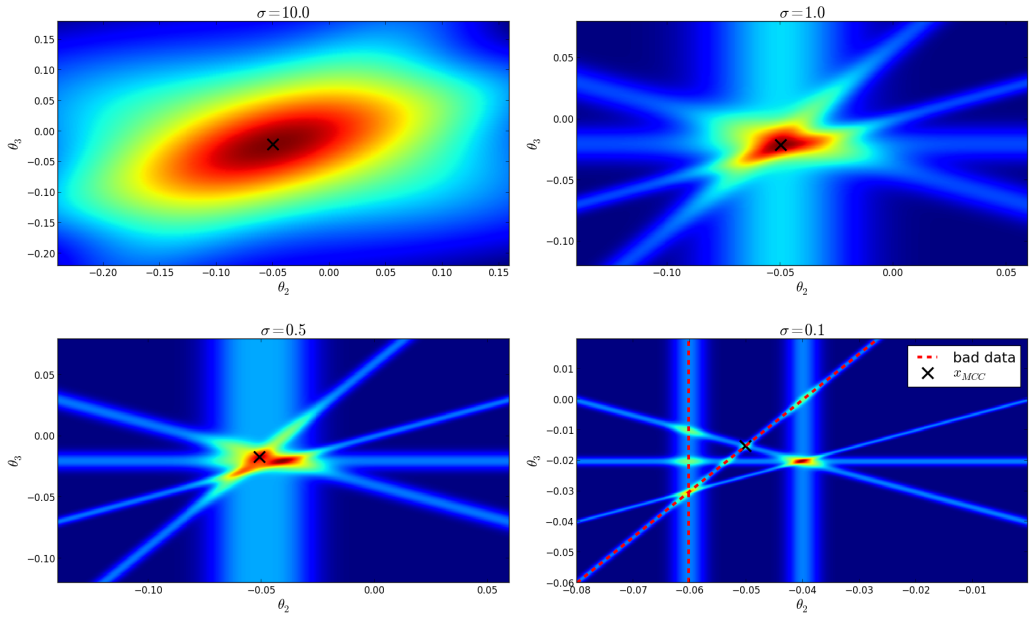


Figure 3.14: 2D projection plot of the Correntropy function with 2 interacting and conforming GE $(-2, 1)$ in measurements P_2 and P_{2-4}. The MCC estimation ends in the local maximum and fails to identify bad data.

On the other side, the optimization of the MCC with EPSO (Section 3.2.2) justifies robust features of the Correntropy function and successfully rejects multiple GE in all scenarios, independent on their conformity. Therefore, it can be preliminary concluded that the Gauss-Newton method, coupled with kernel annealing, does not allow for maximum extraction of information in all cases, and results in a misidentification of GE.

In order to perform a more exhaustive test, a 1000 scenarios of 2 GE in the 3-bus system are once again considered, similarly to what has been described in section 3.2.2. The MCC with a Gauss-Newton optimization managed to correctly identify GE in 81.3% of cases. If compared with the results from Table 3.5, the MCC was still slightly less effective then the LNRT. In order to investigate the MCC performances in more detail, more exhaustive tests on the 4-bus system case study are performed in the following section.

3.3.2 Case Study - the 4-bus AC System

In this section, the proposed MCC estimation procedure is tested on the 4-bus AC system [128]. In order to derive a statistically sound conclusion, each analysis considers 1000 measurement

scenarios. One-line diagram of the system, with location of voltage and active/reactive power meters is shown in Figure 3.15. Each measurement scenario was synthetically generated from the solution of a converged power flow. Afterwards, the Gaussian-distributed random errors were added on the power flow solution in order to mimic the real measurements. All details about the system and measurements generation procedure are available in Appendix A.1.

In order to evaluate and compare the estimation accuracy with respect to the true values of state variables, the following metrics were considered (as proposed in [129]):

$$\varepsilon^{V}_{abs,\omega} = \frac{\sum_{i=1}^{N} |\hat{V}_{i,\omega} - V^{true}_{i,\omega}|}{N} \tag{3.44}$$

$$\varepsilon^{\theta}_{abs,\omega} = \frac{\sum_{i=2}^{N} |\hat{\theta}_{i,\omega} - \theta^{true}_{i,\omega}|}{N} \tag{3.45}$$

where:

- $\hat{V}_{i,\omega}$ estimated voltage magnitude for the i^{th} bus in the scenario ω,
- $\hat{\theta}_{i,\omega}$ estimated voltage angle for the i^{th} bus in the scenario ω,
- $V^{true}_{i,\omega}$ voltage magnitude obtained as a solution of the converged power flow for the scenario ω (free of Gaussian error impact),
- $\theta^{true}_{i,\omega}$ voltage angle obtained as a solution of the converged power flow for the scenario ω (free of Gaussian error impact),
- N number of buses in the entire system,
- $\varepsilon^{V}_{abs,\omega}$ average absolute error of the voltage-magnitude estimates regarding the true values for scenario ω,
- $\varepsilon^{\theta}_{abs,\omega}$ average absolute error of the voltage-angle estimates regarding the true values for scenario ω.

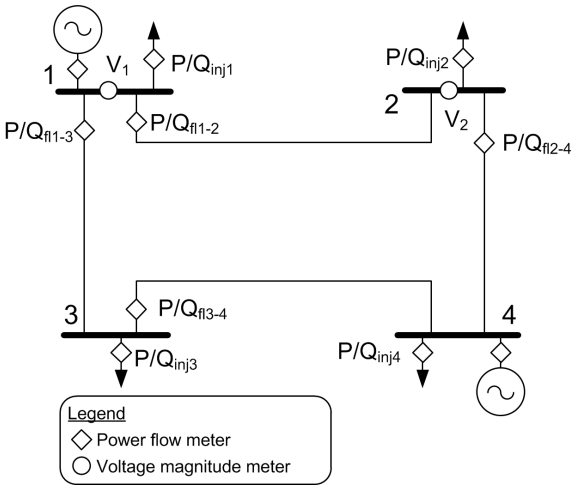


Figure 3.15: Single-line diagram of the 4-bus system [128], with locations of power and voltage meters.

Moreover, in order to quantify the computational requirements of the considered methods, the CPU time t_ω is calculated for each estimation and presented with its statistical properties. In regard to the gross errors, 3 datasets are generated: without the gross errors, with 1 GE per scenario and with 2 GE per scenario. For the purpose of comparability, apart from the MCC estimator, the common WLS estimator and the WLS which includes LNRT iterative scheme for identification and rejection of bad data have also been tested.

The results of estimation accuracy and computational time for the case without any GE are presented in Tables 3.6 and 3.7, respectively. For each average error and computational time, all statistical properties are given: mean, standard deviation, minimum and maximum. It can be observed that estimation accuracy is identical for all of the estimators. The explanation for this is simple: if no bad data are detected, both LNRT and MCC are practically reduced to the WLS method. Specifically, after the first few iterations with large kernel size (where MCC is equivalent to WLS), the MCC estimator will abruptly reduce the kernel size to a minimum value and stop the estimation process. Only LNRT shows a slightly higher computation time, which is due to calculation of sensitivity matrix, the diagonal of which is used to normalize the residuals.

Table 3.6: Estimation accuracy results - 4-bus case study without GE.

	$\text{mean}(\varepsilon^{\theta}_{abs,\omega})$ [rad]	$\text{std}(\varepsilon^{\theta}_{abs})$ [rad]	$\max(\varepsilon^{\theta}_{abs})$ [rad]	$\text{mean}(\varepsilon^{V}_{abs})$ [p.u.]	$\text{std}(\varepsilon^{V}_{abs})$ [p.u.]	$\max(\varepsilon^{V}_{abs})$ [p.u.]
WLS	0.0002	0.0001	0.0008	0.0028	0.0022	0.0152
LNRT	0.0002	0.0001	0.0008	0.0028	0.0022	0.0152
MCC	0.0002	0.0001	0.0008	0.0028	0.0022	0.0152

Table 3.7: Computational performance results - 4-bus case study without GE.

	$\text{mean}(t_\omega)$ [s]	$\text{std}(t_\omega)$ [s]	$\min(t_\omega)$ [s]	$\max(t_\omega)$ [s]
WLS	0.0116	0.0006	0.0110	0.0150
LNRT	0.0116	0.0006	0.0110	0.0160
MCC	0.0116	0.0006	0.0110	0.0150

The following test includes a single gross error per each scenario. Both robust state estimators, the LNRT and MCC are not expected to have any problems with identification of bad measurements. In case of WLS, one expects a certain distortion of the estimation accuracy due to a smearing effect on residuals caused by the GE. These hypotheses are confirmed by the results from Tables 3.8 and 3.9. The average error of WLS estimation is significant, while both the LNRT and MCC show identical performances as in the case without any GE. In regard to the computational expenses, the LNRT and MCC increased the time consumption due to identification and additional estimation needed after the elimination of GE.

Table 3.8: Estimation accuracy results - 4 bus case study with 1 GE.

	mean($\varepsilon^{\theta}_{abs,\omega}$) [rad]	std($\varepsilon^{\theta}_{abs}$) [rad]	max($\varepsilon^{\theta}_{abs}$) [rad]	mean(ε^{V}_{abs}) [p.u.]	std(ε^{V}_{abs}) [p.u.]	max(ε^{V}_{abs}) [p.u.]
WLS	0.0030	0.0021	0.0136	0.0202	0.0248	0.1423
LNRT	0.0002	0.0001	0.0009	0.0028	0.0020	0.0115
MCC	0.0002	0.0001	0.0009	0.0028	0.0020	0.0114

Table 3.9: Computational performance results - 4 bus case study with 1 GE.

	mean(t_ω) [s]	std(t_ω) [s]	min(t_ω) [s]	max(t_ω) [s]
WLS	0.0148	0.0029	0.0110	0.0260
LNRT	0.0269	0.0029	0.0220	0.0380
MCC	0.0261	0.0056	0.0190	0.0920

The most challenging task for robust state estimators is expected in case of multiple gross errors. In this scenario, 2 gross errors are generated randomly in active/reactive power measurements with a value of 20-200 standard deviations of normal measurement errors in 10000 scenarios. In Table 3.10, it can be noted that, unlike the case with a single GE, the estimation distortion has now appeared also in the case of LNRT and the MCC. This is due to the fact that in case of multiple GE, the MCC and LNRT are able to fail in identification of gross errors, which can lead to a significant error of estimated state vector from the true one. The level of the average voltage magnitude and angle errors is again almost equal for both the LNRT and the MCC. In regard to computational times (Table 3.11), the MCC is slightly more efficient since it has no need for a sensitivity matrix evaluation. Expectedly, both the MCC and the LNRT need more time in comparison to the case of a single GE.

Table 3.10: Estimation accuracy results - 4-bus case study with 2 GE.

	mean($\varepsilon^{\theta}_{abs,\omega}$) [rad]	std($\varepsilon^{\theta}_{abs}$) [rad]	max($\varepsilon^{\theta}_{abs}$) [rad]	mean(ε^{V}_{abs}) [p.u.]	std(ε^{V}_{abs}) [p.u.]	max(ε^{V}_{abs}) [p.u.]
WLS	0.0051	0.0047	0.0664	0.0799	0.0934	0.5639
LNRT	0.0003	0.0013	0.0438	0.0028	0.0022	0.0196
MCC	0.0003	0.0014	0.0341	0.0029	0.0023	0.0301

In summary, in the case of a single GE, the MCC estimator always manages to identify and eliminate it. However, in cases with multiple GE, some deterioration appears in estimation accuracy, which is probably caused by GE misidentification. In order to investigate these cases, the efficiency of bad data identification and occurrence of failures is calculated. If the identified GE are not in accordance with true GE, a failure is detected. The results are given in Table 3.12 for the MCC, LNRT and the Largest Residual Test (LRT). The LRT method procedure is identical

Table 3.11: Computational performance results - 4-bus case study with 2 GE.

	mean(t_ω) [s]	std(t_ω) [s]	min(t_ω) [s]	max(t_ω) [s]
WLS	0.0157	0.0030	0.0110	0.0280
LNRT	0.0407	0.0061	0.0230	0.0930
MCC	0.0374	0.0159	0.0220	0.2730

to the LNRT except that it uses residuals which are not normalized. The LNRT is slightly more efficient then the MCC, while the LRT is by far the worse. One can justify the LNRT superiority by normalization of residuals, what is proved to contribute to robustness in regard to bad data identification. In any case, it is significant to note that the MCC is more robust then the LRT. These results suggest that some residual or kernel size normalization should also be adopted to MCC to improve the robustness.

Table 3.12: Efficiency in 2 GE identification of LNRT and MCC considering 1000 measurement scenarios of the 4-bus AC system.

	Wrong	**Correct**	**Efficiency**
LRT	381	9019	90.19%
LNRT	166	9834	98.34%
MCC	300	9700	97.00%

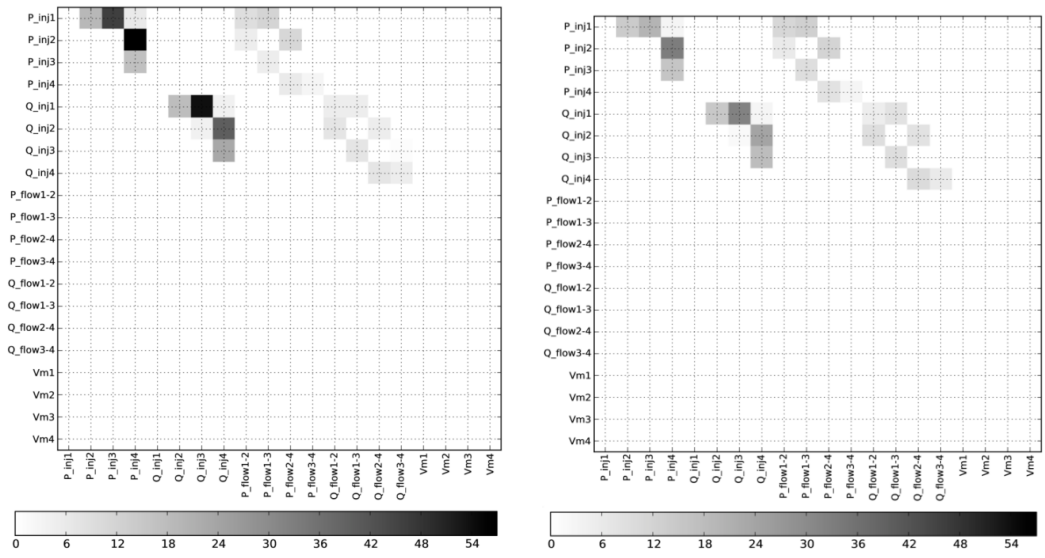
The frequency of particular failure-causing GE combinations are comparatively depicted in Figure 3.17 for LRT and MCC and in Figure 3.16 for LNRT and MCC. The following conclusions, which are of key importance for better understanding of MCC properties, can be derived from these results:

- ▷ There is a high coincidence between the cases when MCC, LRT and LNRT fail to identify bad data,
- ▷ The LNRT is slightly more robust than MCC,
- ▷ The MCC is found to be more efficient then the LRT,
- ▷ GE scenarios with the highest frequency of failure are the ones that include incident bad data,
- ▷ Methods never fail in cases where both GE come from measurements of power flows,
- ▷ At least one of GE needs to come from the injection power measurement in order to cause a failure,
- ▷ Failures occur exclusively in cases when both GE come from active or reactive power measurement,

▷ There are no failures which include GE in voltage magnitude measurements.

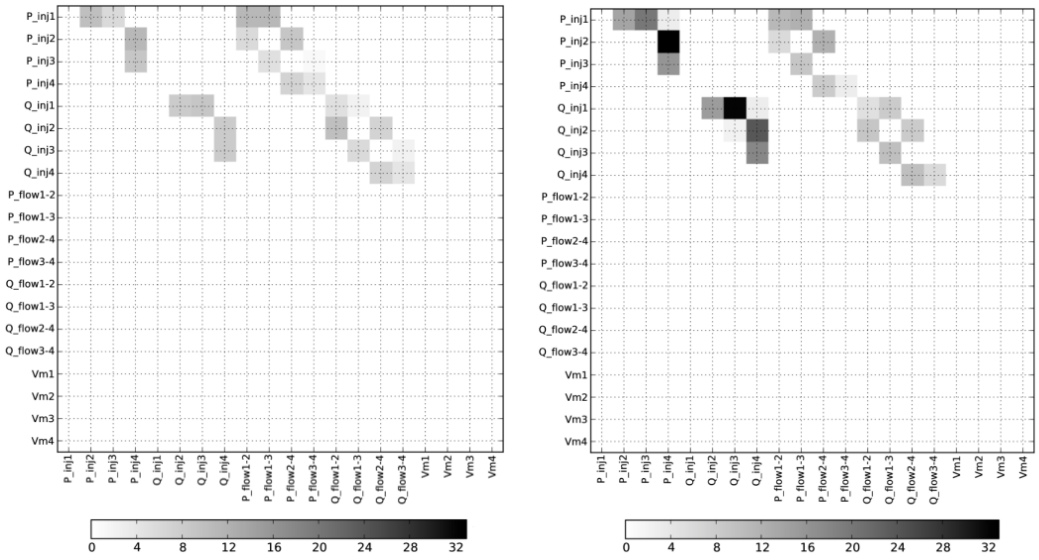
These results point to a high similarity between the LNRT (LRT) and MCC state estimation in regard to their robust performances. If one looks in more detail at the proposed Newton's method and kernel annealing in the MCC estimator, a strong similarity to the elimination of largest residuals can be observed. The initial iterations, prior to GE detection, remain the same since MCC with large kernel size corresponds to the WLS. Afterwards, the Parzen window is annealed according to the largest residual. Unlike the LRT or LNRT, a measurement which corresponds to the largest residual is not automatically eliminated, instead, the Gauss-Newton step is the one to determine the destiny of bad data associated with the largest residual. Also, MCC does not eliminate and rerun the estimation, it only annuls the impact of the identified bad data on the estimation result.

Finally, it can be concluded that the proposed Newton's method and kernel annealing strategy do not allow maximum extraction of information in all the scenarios, since it leaves room for the masking effect which is able to mislead the Newton's search to the local optimum of the Correntropy function.



(a) LRT (b) MCC

Figure 3.16: The occurrence of failures in identification of 2 randomly generated GE in 10000 load scenarios of the 4-bus system for LRT (left) and MCC (right).



(a) LNRT (b) MCC

Figure 3.17: The occurrence of failures in identification of 2 randomly generated GE in 10000 load scenarios of the 4-bus system for LNRT (left) and MCC (right).

3.3.3 Case Study - the 24-bus IEEE RTS System

Robust performances of the MCC state estimator in identification and elimination of bad data are once more preliminary tested and confirmed for the system of larger scale - the IEEE RTS 24-bus [130]. All details about the network are available in Appendix A.2.

Since single bad data do not present any threat to the MCC SE, a particular attention is paid to multiple GE. When performing a statistically sound analysis, all possible bad data scenarios need to be simulated by generating 2 gross errors in all possible couples of available measurements. Such a Monte Carlo simulation would require consideration of $\binom{m}{2}$ bad data combinations for a sufficient number of GE values. This would lead to an insuperable number of estimations needed to obtain some statistically meaningful analysis.

Therefore, the use of *Importance Sampling* method is suggested. It distorts the PDF (and Cumulative Density Function (CDF)), during which it increases the probability of important events at the cost of irrelevant ones. More information about the importance sampling method can be found in Appendix C.2, according to [131]. The important events in robust power system state estimation are failures of the estimator to identify bad data, i.e. the events of failure. According to the results for SE of the 3-bus (section 3.3.1) and 4-bus (section 3.3.2) systems, it is quite straightforward to characterize the scenarios in which the probability of failure is high as well as the GE scenarios which are always identifiable. In principle, failures are most probable to occur under the following assumptions:

- Multiple interacting (and conforming) gross errors,
- One of the GE occurs in the injection power measurement,
- At least two interacting GE are in active (reactive) power measurements.

Since the probability of failure is substantially lower than the probability of success, the purpose of importance sampling is to increase sampling of cases probable to cause failure, at the cost of cases which will certainly be identified. Therefore, in order to keep the expected value, it is necessary to compensate this distribution distortion by changing the indicator function in a reverse sense.

In a particular case study, when precision of $\beta = 5\%$ was assumed, it resulted in $N = 3760$ simulated samples where gross errors were introduced only in active power measurements. Thus, the sample size was substantially reduced thanks to the importance of sampling method which enabled a significant convergence acceleration of the method.
The results are illustrated in Figure 3.18 which indicates the combinations of 2 GE which resulted in failures of the MCC and LNRT methods. The blue spots indicate couples of measurements the sampling of which was favored. Since these important measurement couples were defined due to their incidence, the blue spots depicted in Figure 3.18 also represent the sparsity of the Jacobian matrix.

The results confirm the same properties of the MCC estimator that were revealed through the tests on the 4-bus system (Section 3.3.2):

◇ Failures are probable to occur only if the GE are incident,

◇ There is a high coincidence between failures of the MCC and LNRT methods,

◇ Failure is probable to occur only if one of bad data is injection power,

◇ Failure is probable to occur only if both bad data are active or reactive power,

◇ Identification efficiency of LNRT is still slightly beyond the MCC estimator.

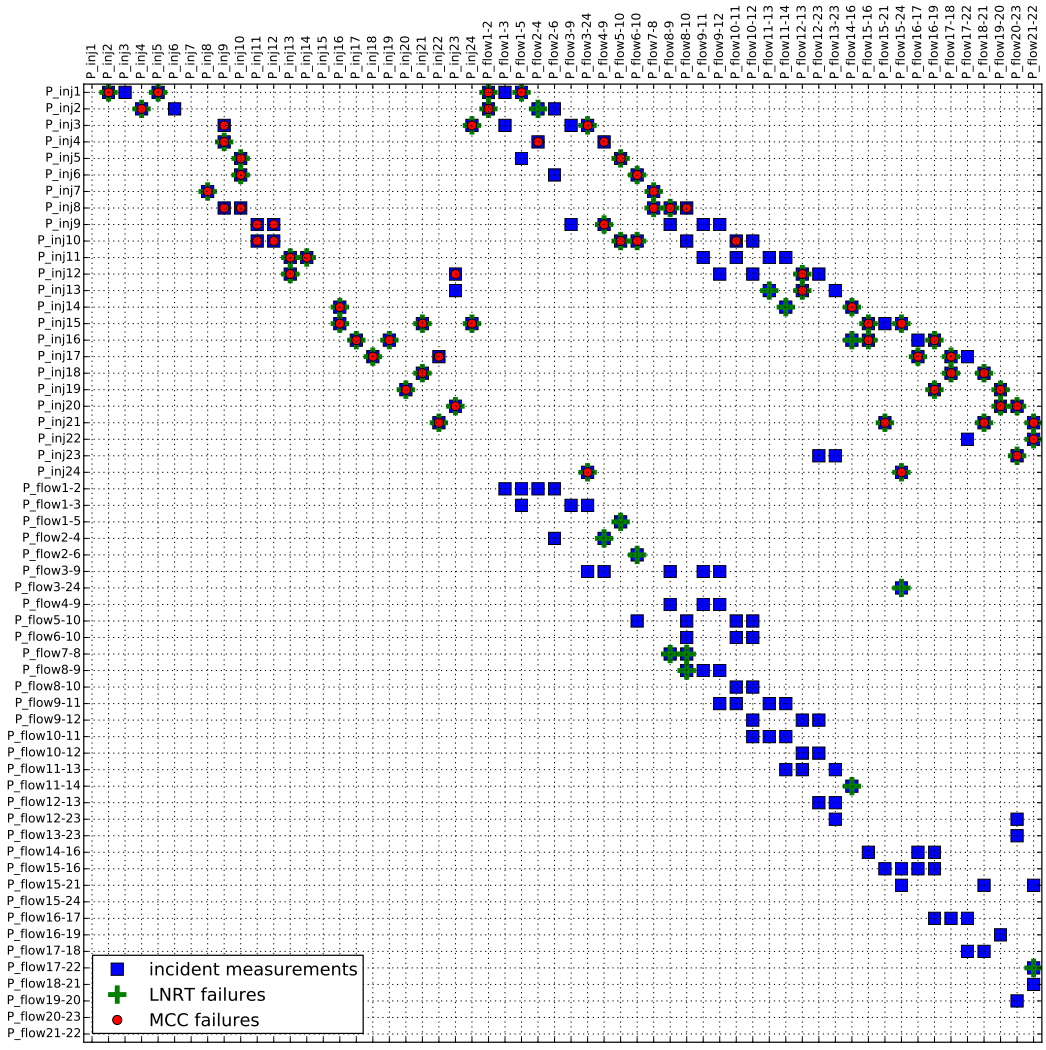


Figure 3.18: The occurrence of MCC and LNRT failures in GE identification. A high coincidence between the MCC and LNRT failure scenarios can be noted. Also, a failure is most probable to occur when GE are incident and at least one is in the active power measurement. Failure probability is zero if one of the GE is in active and the other in reactive measurement.

3.3.4 Conclusions Derived - Problems of the MCC State Estimator

From the considered case studies, some key important conclusions can be derived about the properties of the proposed state estimation based on Correntropy. In the preliminary tests of the toy problem (3-bus DC system), the MCC was optimized by the meta-heuristic method EPSO. The results indicate a successful identification of all considered scenarios for multiple GE thus outperforming the conventional robust estimators which are prone to failure in certain cases. The key significance of these results is in justification of the proposed criterion. The global maximum of the Correntropy function always corresponds to the true solution which is not deteriorated by erroneous data and simply ignores their presence. The other local optima, which emerge with a small enough kernel size, correspond to misidentification, i.e. consider bad data as correct, while some of the true measurements are suspected to be erroneous. One can say that the EPSO method always obtains result on the global maximum, i.e. extracts maximum of information from the acquired measurement dataset. The only case in which global maximum of the Correntropy function is expected to be biased from the true solution is when a critical k-tuple of measurements is infected by GE. It is important to emphasize that any other static state estimator is also helpless in that scenario.

Unfortunately, since the EPSO cannot be industrially-acceptable solution, the Newton's method is adopted as an alternative in order to allow for feasibility in real time environment. Due to complexity of the correntropy function, Newton's search is coupled with an adaptive kernel annealing strategy that ensures required smoothness and concavity in any iteration step. This results in a fast-converging estimation method which performs all the bad data processing functions (detection, identification and elimination) simultaneously and in most of the cases successfully identifies multiple bad data.

It is of particular interest to observe the novel mechanism of GE detection. The method is initiated with the large kernel size, or equivalently with the WLS solution. In case this solution does not change through the annealing process, i.e. if the Newton's step size is insignificant from largest to the lowest value of σ, one can deduce that the available measurements vector does not contain any GE. In that case, the MCC will result with the state vector equivalent as the one provided by the WLS. Otherwise, if the measurements set is corrupted by the GE, the solution will keep deviating from the initial WLS solution, until their deteriorating effect is eliminated.

Moreover, the process of bad data elimination by the MCC estimator is specific and different in nature, in comparison to elimination by the largest residual test. Unlike the LNRT, the MCC in fact does not literally eliminate bad measurements from the dataset, it only annuls their contribution in the objective function, i.e. neutralizes their impact on the solution. The results are in most cases the same, but the key difference is that the MCC does not require any restructure of the relevant matrices or any re-processing.

However, through all the exhaustive test case studies, certain scenarios have been detected when the proposed MCC estimator misidentifies bad data and provides a corrupted estimation solution. In other words, the iterative solution ends up in a concave region which finally emerges

as a local optimum. Therefore, one can conclude that the adopted optimization procedure based on Newton's method and Parzen window annealing, is unable to extract the information maximum that is set by the Correntropy criterion in all the cases. The initiation with the large kernel size equates MCC to WLS and thus leaves room for the well-known masking effect among residuals, caused by multiple interacting GE. These problematic GE scenarios are relatively low probable, but well-known in the robust regression theory since are able to harm any estimator that maximizes some function of residuals. They can be characterized as the following:

- Multiple interacting and conforming GE,
- At least one of bad data corresponds to a Leverage Point.

Hereafter, a more exhaustive discussion about the impact of the problematic GE variants on the state estimation procedure is provided. Moreover, one recognizes strong relationships between the Correntropy and the non-quadratic criteria through the link of M-estimation.

3.3.4.1 Conforming gross errors

It can be recognized from the test cases considered in this chapter, that the probability of MCC estimator to misidentify bad data is significantly elevated when the GE are interacting. Two (or multiple) gross errors are interacting when they affect measurements whose errors mutually impact their estimative, i.e. error in one measurement impacts the residual of the other measurement and vice-versa. Logically, interaction is strongest for the incident measurements.
If GE are interacting, one can ask about the values of GE that affect particular measurements. Regarding this, the worse case scenario is when GE assumes values that support each other's residuals. In that case they are addressed as conforming GE. Therefore, if multiple GE are interacting and possibly conforming, they represent a very serious threat to any estimator that optimizes some function of residuals. They cause the so-called masking effect among the residuals. The interaction impact between the bad data adds up and makes the residuals of good measurements larger then the residuals of bad data. This will cause, in most of the known robust state estimators, a misidentification of GE, i.e. an estimated state of a system which is far from the true one. A very illustrative sample that demonstrates how conformity of GE can harm the MCC state estimator is shown for the 3-bus DC system in Figures 3.13 and 3.14.

Moreover, some characteristic GE scenarios of the 4-bus AC system case study can be observed. As shown in section 3.3.2 GE are misidentified only in cases when they are incident. Now, one can examine which of the incident GE are particularly dangerous. A couple of measurements P_3 and P_{1-3} are chosen and 1000 scenarios of these 2 GE were randomly generated with uniform distribution in the range $[-1.0,-0.1]\cup[0.1,1.0]$. This way, GE bins are generated in all 4 quadrants. Figure 3.19 shows a scatter-plot for all generated GE (gray points). Among these cases, the cases when MCC estimator failed in identification are highlighted in red. It is evident that failures occur exclusively in the narrow area from 2^{nd} and 3^{rd} quadrant. These particular GE values appear in measurements P_3 and P_{1-3} as conforming and cause a masking effect which

misleads the Newton's method through σ-annealing to the local maximum of Correntropy function instead of the global.

It is necessary to emphasize that the same masking effect of multiple conforming GE is a well-known problem in the robust state estimation theory. It is prone to misleading any estimator that optimizes a function of residuals like: LNRT, WLAV, Non-quadratic criteria, or any other M-estimators.

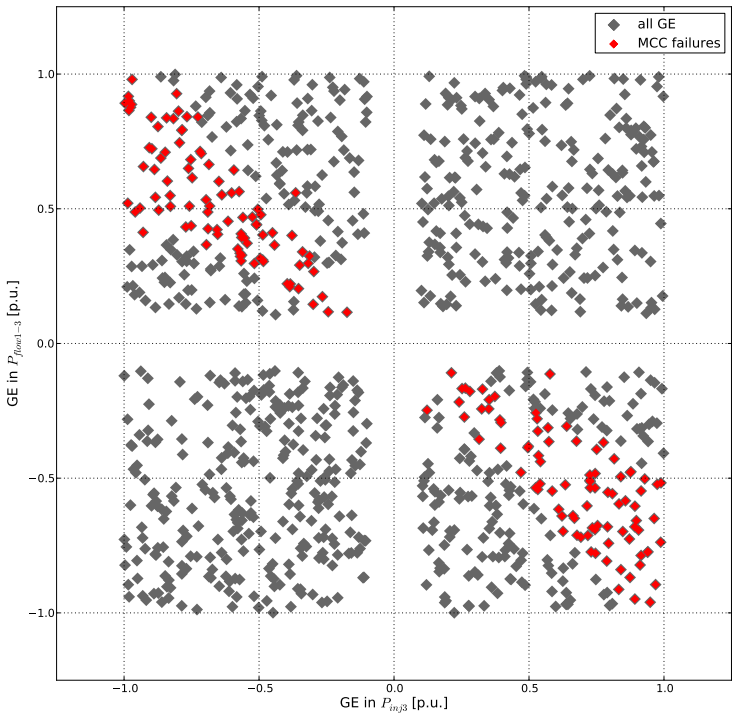


Figure 3.19: Scatter plot of 2 GE in P_3 and P_{1-3} for 1000 scenarios. Out of all GE (gray points) generated in the range $[-1.0,-0.1]\cup[0.1,1.0]$ only conforming cases cause failure of the MCC estimator.

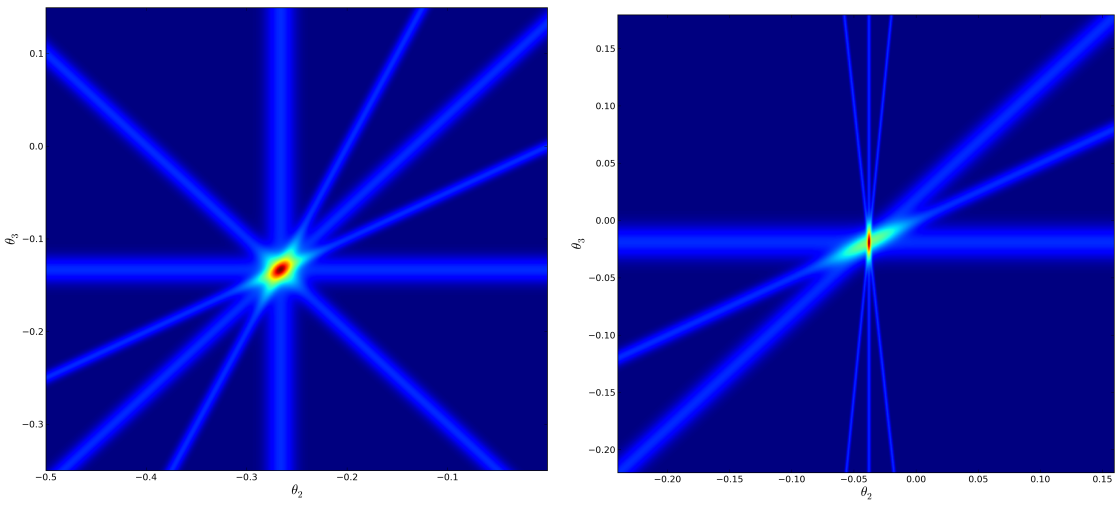
3.3.4.2 Leverage points

When investigating the cases of MCC failures in the 4-bus and 24-bus systems, it can be observed that no failure is recorded when both GE are in power flow measurements. In other words, at least one of GE is required to be in an injection power measurement in order to harm the MCC estimator. A reason for this comes from the fact that all measurements do not have the same impact on the estimation result. Specifically, this impact will particularly depend on the structure of the corresponding equation, i.e. on many factors like: measurement type (injection, flow, active power, reactive power, voltage), its position in the system, local redundancy and relevant network parameters. In that sense, measurements which will have an expressed influence on their estimate are recognized as the so-called Leverage Points. These measurements make an outlier in the factor

space defined by rows of Jacobian matrix and thus have an undue influence on the state estimate. Therefore, a measurement residual that corresponds to a leverage point will be relatively small even when it is contaminated by a large error. Accordingly, such a bad leverage point is very difficult to be identified by methods which rely on residuals [4]. One can say that an end-case of a leverage points would be a bad critical measurement whose residual is always equal to zero. However, a residual of a leverage point is never zero and unlike the critical measurement it can be removed without the loss of system observability.
The following types of measurements are probable to behave like the leverage points:

- An injection measurement placed at a bus which is incident to a large number of buses,
- An injection measurement placed at a bus which is incident to a branch with an impedance value significantly lower then the other branches,
- Flow measurement along a branch whose impedances are significantly lower from those of the other branches in the system.

The above-mentioned conditions in a power system will affect the numerical structure of the measurement Jacobian H and generate outliers in the corresponding factor space, i.e. leverage points. The rows of the Jacobian that correspond to leverage measurements will have entries of very different magnitudes compared to those of the other rows. A leverage point could also be created artificially (e.g. in WLS or WLAV) along the way so that a large weight is attributed to a specific measurement.



(a) w/o Leverage points (b) w/ Leverage points

Figure 3.20: Visualization of a leverage point impact on the J_{MCC} objective function ($\sigma = 0.1$). Left: objective function without the Leverage point (all lines have equal impedance). Right: line 1-2 is shortened 10 times thus creating leverage points in measurements P_{1-2}, P_1 and P_2).

The Correntropy criterion is equivalent to maximization of entropy of residuals around zero. Since the MCC estimator is initiated by the WLS, it is also expected to be sensitive to the leverage

point effect. One of the most common leverage points in a power system are injection power measurements, especially if they are incident with many branches. Finally, it is now clear why the MCC failed to identify the GE only in cases when one of bad data was an injection power.

In order to visualize the impact of leveraging points on the Correntropy function, the 3-bus DC system from Figure 3.3 has once again been used, but with 2 different settings of the parameters: one sample considers equal impedances in all 3 lines and the other one assumes the line 1-2 as 10 times shorter. Figure 3.20(a) represents the J_{MCC} objective function of the 3 bus DC system which does not have any expressed leverage point since its topology and parameters are set as symmetrical and equivalent. Figure 3.20(b) corresponds to the same system but considers the line 1-2 which is 10 times shorter. It is significant to note a big change in the objective function caused by the change of a single parameter which created leverage points in the measurements P_{1-2}, P_1 and P_2.

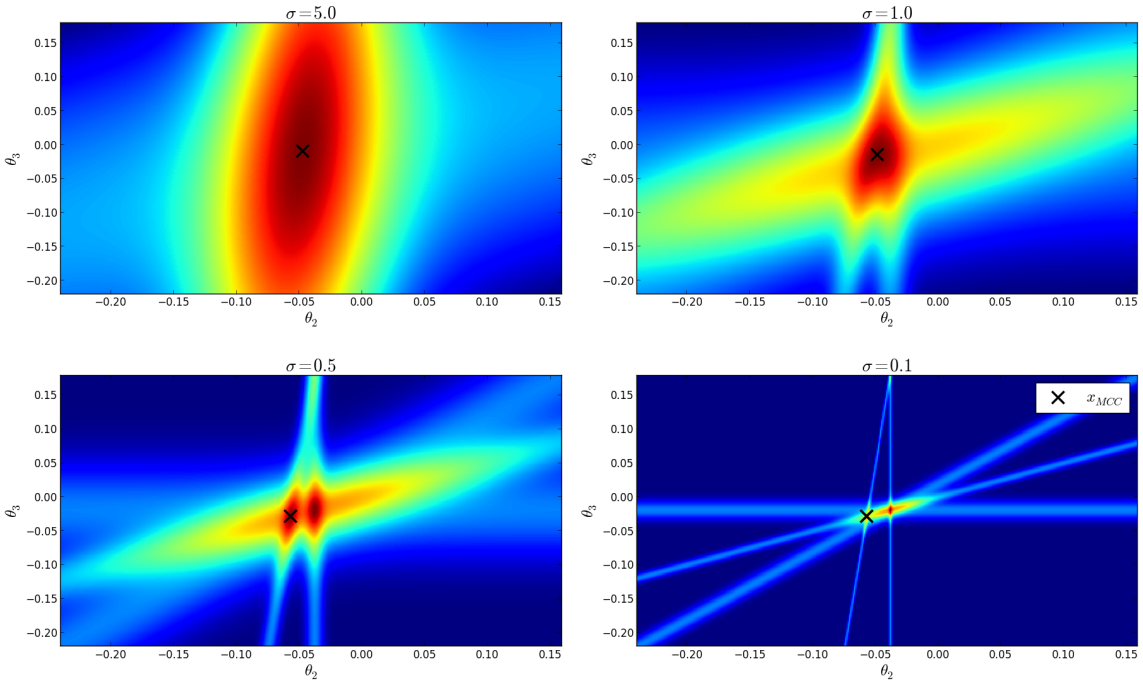
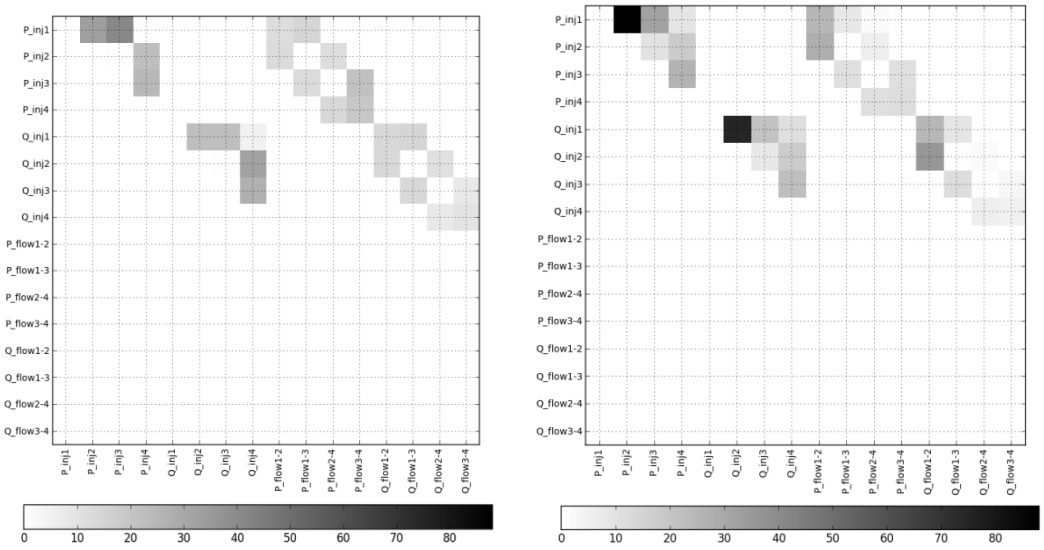


Figure 3.21: 2D projection plot of the Correntropy function for the 3-bus DC system with leverage points, for kernel sizes $\sigma = \{5, 1, 0.5, 0.1\}$. A bad leverage point in P_1 has a strong influence on the objective function shape and manages to attract the solution of the Gauss-Newton while the Parzen window is annealing. At the end, the bad leverage point is identified as a true measurement at the cost of the other 3 measurements which are correct.

If some of these leverage points contain GE, the MCC estimator will have a difficulty in identifying it. Figure 3.21 illustrates a sample of how even a single bad leverage point in P_{inj1} can cause an MCC estimation failure. In the phase when Parzen window is large, leverage point attracts the solution toward itself and reduces the corresponding residual. When σ shrinks and the local optima start to emerge, the solution remains trapped within the local optima. Finally, the bad

leverage measurement is not rejected but the 3 other correct measurements are rejected instead. In any case, the problem of a single bad leverage point is easy solvable by normalization of Parzen windows, as proposed in section 3.4. More serious concern is when one of the bad data is the bad leverage point.



(a) Occurence of MCC failures - without Leverage points

(b) Occurence of MCC failures - without Leverage points

Figure 3.22: Intensity of MCC failure occurrence for the 4-bus system with and without the leverage points. If the left and right figures are compared, it can be noted that leverage points, when they contain GE, cause much more failures then the common measurements.

Furthermore, the impact of a leveraging effect on the MCC state estimation is again investigated on the 4-bus case study. In the original case described in section 3.3.2 (Figure 3.15) the system has 4 buses and 4 lines, where all the buses are incident with 2 lines and impedances of all the lines which are equal to $Z = 0.01 + j0.05\Omega$. Accordingly, since the system and meters are set completely symmetrically, there were no an expressed leveraging effect. In order to introduce a leverage point, impedance of the line 1-2 was decreased 10 times. The simulation is performed with random generation of 2 GE in power measurements in order to reconstruct the level of leveraging impact. The occurrence of MCC failures is shown in Figure 3.22 for the case with and without the leverage points. It is important to observe the high frequency of failures in the cases when GE are present in the measurements P_1, P_2 and P_{1-2} and also in the reactive measurements Q_1, Q_2 and Q_{1-2}. Therefore, the failure of the MCC is much more probable to occur when GE are located in the leverage points. In conclusion, leverage points are in fact not undesirable at all, unless they contain GE, since the good leverage points contribute to estimation robustness and reliability.

The effect of leverage points was early recognized and attempted to be treated through years of research work in order to improve the robust features of state estimation in power

systems. The first attempt to regulate the leveraging effect was the proposal of residual normalization [1]. More specifically, the diagonal of the hat matrix (or sensitivity matrix) defined as $K = H(H^T R^{-1} H)^{-1} H^T R^{-1}$ (where H is Jacobian matrix), should represent the sensitivity of the measurement residuals to the measurement errors. Thus, if diagonal entry K_{ii} is close to 1.0, the i^{th} measurement is likely to behave as a leverage point. Geometrically, K_{ii} provides a measure of distance of the vector of Jacobian matrix row H_i, corresponding to the i^{th} measurement, from the bulk of the remaining measurement factors [4]. The matrix K permits calculation of the residual covariance matrix as $\Omega = (I - K)R$, whose diagonal entries are used to normalize the residuals and the off-diagonal elements should indicate the interaction level among measurements.

It has been early recognized that M-estimators, i.e. any estimator that minimizes the sum of certain functions of residuals, are sensitive to bad leverage points. In order to enhance the robustness, the idea was (the same as in LNRT) to replace residuals with normalized residuals which resulted in so-called generalized M-estimators (GM-estimator), as proposed in [9]. This methodology is proven to work in the case of a single bad leverage point but it is still prone to a masking effect in the presence of multiple leverage points. Also, it is important to know that down-weighting of all leverage points is not the best solution since good leverage points can significantly improve estimation reliability. The work published in [21] proposed the Schweppe-Huber generalized M-estimator that selectively down-weights bad leverage points using weights calculated from projection statistics, which can deal with multiple outliers. Among the M-estimators, it is also significant to observe the weighted least absolute value (WLAV). It has been proven to be very efficient in rejection of bad data as long as they are not leverage points. In that sense, Çelik and Abur in [19] proposed a linear transformation which modifies measurement equations in order to remove leverage points and apply the WLAV estimator in order to automatically reject bad data. In the same sense, a generalized MCC estimator is proposed in section 3.4.

The other group of robust state estimators, immune to leveraging effect, are the so-called high-breakdown point state estimators. The name comes from the fact that such estimators are able to provide a correct solution even if half of the redundant measurements are bad leverage points. The main representatives are Least Median of Squares (LMS) and Least Trimmed Squares (LTS) estimators, based on the methods initially developed by Rousseeuw and Huber in [22] and adapted for an electric power system by Mili *et al.* in [26] and [27]. Despite remarkable results regarding robustness to bad data, these methods are commonly limited by high computational burdens, barely acceptable for real-time SE. A more exhaustive discussion about these estimators, can be found in section 2.2.1.

3.3.4.3 MCC and M-estimation

With the establishment of the Correntropy function, it has been recognized in [132] that it belongs to the M-estimation framework by its definition. Accordingly, this link is used here in order to discuss the relationship of the proposed MCC estimator and the conventional non-quadratic state estimation.

The earliest attempts to enhance the robustness of the WLS estimator have resulted in various types of the so-called ***Non-Quadratic Criteria*** (**NQC**). The principal idea was to eliminate the high contribution of the outliers in the objective function on the way to change the quadratic function to some other form that is increasing less steeply after a residual exceeds a certain value. The most known representative objective functions are quadratic-linear (QL), quadratic-tangent (QT), quadratic-square root (QR) and quadratic-constant (QC). They are all defined when the quadratic function changes to linear, tangent, square-root or constant function, respectively, when the absolute value of a residual is larger then a predefined threshold T. The intentions of "breaking" the quadratic function is generally the following: assign smaller or insignificant importance to larger errors, while keeping the least squares criterion for the small errors. The relevant literature about the proposal and development of these criteria for robust state estimation, is available in [8], [9], [11], [12], [18].

All the non-quadratic criteria can be set under the same framework called the M-estimation, proposed by Huber in [22] as a systematic way of handling outliers and introducing the concept of robustness in maximum likelihood theory. Generally, the M-estimator is defined as:

$$\min_x \sum_{i=1}^{N} \rho(r_i) \tag{3.46}$$

or

$$\min_x \sum_{i=1}^{N} \psi(r_i), \quad \text{where } \psi(x) = \frac{d\rho(x)}{dx} \tag{3.47}$$

Thereby the $\rho(x)$ must obey the following properties:

1. $\rho(x) \geq 0$
2. $\rho(0) = 0$
3. $\rho(x) = \rho(-x)$
4. $\rho(x_i) \geq \rho(x_j), \quad |x_i| > |x_j|$

If one can recall the definitions of the ITL criteria, like the Minimum Error Entropy (MEE) or the Maximum Correntropy Criterion (MCC), one can easily note strong relations with M-estimators. This was recognized and mathematically proven in [132]. Hence, if one puts the MCC in the general framework of M-estimation, the function $\rho(r)$ can be defined as:

$$\rho(r) = \frac{1 - e^{(-r^2/2\sigma^2)}}{\sqrt{2\pi}\sigma} \tag{3.48}$$

Therefore, it is easy to check that function $\rho(r)$ from (3.48) satisfies all the properties of a M-estimator. Moreover, according to 3.48 it can be observed that the Correntropy criterion is the most similar with the so-called Welsch M-estimator [22], defined with the following objective function: $\rho(r) = c^2/2\left[1 - e^{-(x/c^2)}\right]$. However, the proposal for a Correntropy criterion is more powerful as it puts the approach of the Welsch M-estimator into a new theoretical perspective. Firstly, it is no longer viewed as a mathematical construct adopted by engineering to achieve

results that "make sense" - the Correntropy concept has a specific meaning and provides a general interpretation, in terms of Information Theory, of the search for similarity between the pattern of the observed values and the patterns allowed by the power flow equations. Secondly, this interpretation opens new avenues for developing solving algorithms, as this thesis will show. And thirdly, it allows establishing a bridge to further development associated with Information Entropy concepts and estimation, an avenue open to an unchartered territory for which only a glimpse is given in Chapter 6.

If the MCC is compared to the non-quadratic criteria, the most relevant is definitely the ***Quadratic-Constant*** (**QC**) criterion. The QC is proposed as an attempt to introduce a threshold of indifference, with a quadratic function for each error below a certain threshold and a flat function beyond this value. Among the other non-quadratic criteria, the QC provides the best suppression behavior, since it most efficiently eliminates those measurements whose residuals are larger then the threshold [12], [18]. It is significant to mention a very recent PSSE comparative study from [129], which reports that QC outperforms other estimators due to computational requirements and estimation accuracy. Generally, the QC state estimator can be defined as:

$$\min_{x} J_{QC}(x) = \sum_{i=1}^{m} s_i(x) \tag{3.49}$$

$$s_i(x) = \begin{cases} r_i^2(x) & \text{if } |r_i(x)| \leq T \\ T^2 & \text{if } |r_i(x)| \geq T \end{cases}$$

The objective function, defined as such, behaves as a quadratic function (WLS) for the residuals r_i, such that $r_i \in [-T, T]$. Otherwise, for residuals outside these boundaries $r_i \in \langle -\infty, T \rangle \cup \langle T, \infty \rangle$ the QC is indifferent to distance, meaning that the impact of that measurement is simply eliminated from the solution. Among these properties, one cannot hide the strong similarity of QC estimator with the MCC estimator, proposed in this thesis. For the purpose of visualization, Figure 3.23 shows all the non-quadratic criteria and the Correntropy function (inverted and translated in order to be more easily comparable, as in eq. (3.48)).

In addition to many similarities, there are also some key important differences, i.e. contributions brought by the Correntropy criterion. First, lets see how QC cost function can be written in mathematical programming form:

$$J_{QC}(x) = \sum_{i=1}^{m} \left[(1 - b_i) r_i^2(x) + b_i T^2 \right] \tag{3.50}$$

where $b_i = \{0, 1\}, \forall i \in \{1, \ldots, m\}$. Hence b is an auxiliary binary variable vector, which makes such a definition a mixed integer non-linear problem (MINLP).

After all, one defines the similarities and differences between the proposed MCC criterion and the QC, as the most relevant among the conventional non-quadratic criteria. First of all, the MCC and QC have the same intention of enabling bad data rejection abilities by breaking the quadratic function to a constant value. Also, they both keep the quadratic core for the errors,

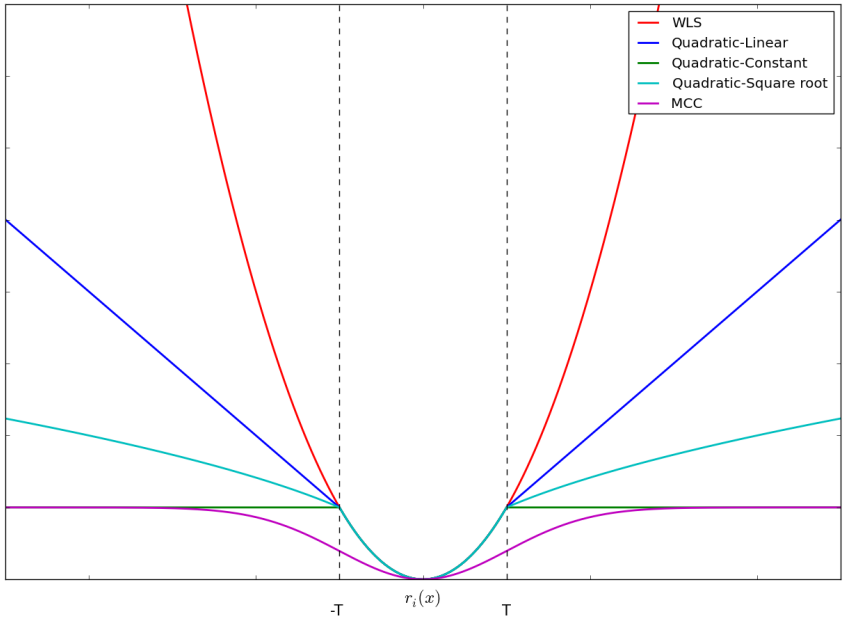


Figure 3.23: A comparative plot of the non-quadratic criteria and the Correntropy function.

smaller than a predefined value. However, a significant difference in that sense is that Correntropy offers much smoother transition from quadratic, through linear and finally to constant function, while the QC abruptly changes directly from the quadratic to constant function. Due to this, QC objective function is non-differentiable at the bounding points and therefore leaves a possibility for occurrence of some non-differentiable local optima. On the other hand, the Correntropy is continuous and differentiable on all of its domain. Also, the MCC does not need any auxiliary binary variables, so it is classified as a common non-linear problem (NLP).

Moreover, the MCC estimator by definition contains a free parameter σ, i.e. size of the Parzen window which manipulates the objective function surface and acts as a tuning parameter for sensitivity on outliers elimination. From the perspective of non-quadratic criteria, an analogy can be found in the break even point T. The non-quadratic criteria, as initially proposed in [8], [9], was established with a predefined fixed value of the threshold. Later, Lo *et al.* [133] recognized the necessity to variate the threshold and proposed first few iterations with the WLS and then switched to QC with a variable breaking point. This concept of shrinking the breaking point with a predetermined starting value and a fixed step length was obviously a substantial improvement comparing to the fixed breakpoint approach. Finally in [10], one demonstrated that this smooth shrinking can barely accommodate all the cases and uncertainties and therefore suggested a strategy that varies the break-even point in order to track the upper bound of the normalized residues of good data.

A very similar strategy is proposed in this thesis for annealing the Parzen window in order to ensure smoothness and concavity in the area of iterative Newton's step. More specifically, due to complexity of the correntropy function (multi-modality) and requirements of the real

time environment, the Newton's method is proposed, coupled with the smooth kernel annealing strategy. Unfortunately, this method is not able to extract the maximum information from the conveyed measurements, i.e. it may fail and misidentify the GE.

Finally, one must emphasize that all the methods established on the non-quadratic criteria are more like algorithmic attempts, with lack of theoretical background that would justify such modeling and still remain tightly connected to the minimum square error. On the other hand, the interpretation proposed in this thesis, under the light of Information Theory, gives a strong theoretical background and justification to all these previous attempts. More than that, it offers a novel perspective on the state estimation problem established on information theory and thus sets a novel reference, based on an information measure, whose maximum always corresponds to the true value.

3.4 Generalized MCC Estimator

In regard to observed inefficiencies in GE identification, this section intends to find room for improvement for the proposed MCC criterion and avail it in order to enhance the robustness to multiple GE.

In tests performed in section 3.3, the parameter σ_i, which defines the size of the Parzen window for the i^{th} measurement, was considered unique for all the available measurements ($\sigma_1 = \sigma_2 = \ldots = \sigma_m = \sigma$). Here, an opportunity for improvement can be observed since σ plays a key role in changing the global shape of the objective function, i.e. the final state estimation solution. Moreover, it is reasonable to assume that the size of the Parzen widow should be adjusted for each particular measurement, depending on its characteristics. As discussed earlier, measurements have different impact on estimated state vector, i.e. on their own estimative. This impact will directly depend on measurement type, location in the system, local redundancy, relevant network parameters etc. The measurements which are more powerful and effective in the estimation process are called leverage points. Therefore, the idea is to regulate the Parzen window of each measurement according to its importance in the estimation process.

The same idea was followed by [1] when proposing normalization of residuals in order to obtain zero mean random variables with variance equal to 1. More specifically, the residuals are normalized by the corresponding diagonal entries in the residual covariance matrix $\Omega = R - H(H^T R^{-1} H)^{-1} H^T$. If the diagonal elements of Ω are considered to be proportional to sensitivity of a measurement to its residual, leverage points would be characterized as low sensitive. In the end-case of critical measurement, the estimated value is determined by its measured value, i.e. the corresponding diagonal entry of the hat matrix is equal to 1. This solution for the elimination of leverage effect is proved to contribute to robustness. However, when affected by multiple GE, a failure is still possible due to the masking effect.

In any respect, the impact of Parzen window normalization is expected to contribute to robustness of the MCC. Specifically, if the Parzen window size σ_i is increased, the contribution of the corresponding measurement to the final result will be decreased and vice-versa, similarly to

the weights in the WLS. This benefit has not been availed until now and all the Parzen windows were set with the same bandwidth. Since all the kernel sizes σ_i need to keep annealing throughout the estimation process, a scaling vector to keep a constant ratio between them is suggested. Hence, the vector δ is introduced in order to scale the kernel size for each measurement in the vector σ. Therefore, each σ_i is now determined considering the δ_i and the maximum residual of true measurements. This leads to the definition of ***Generalized MCC*** (**GMCC**) state estimator, which defines each kernel bandwidth σ_i in equation (3.30) in the following way:

$$\sigma_i = \delta_i \cdot \max_j \left(\frac{z_j - h_j(x)}{\delta_j} \right), \text{ and } i, j \in \{1, \ldots, m\} \backslash \{S\} \tag{3.51}$$

where S represents a group measurements suspected as bad data. This way, the Parzen windows are normalized instead of normalizing the residuals. The question which still remains unanswered is how to define the scaling vector δ. Since there is still no unique, optimal solution, two of them are proposed and an appropriate comparative study will be performed:

- **Normalized MCC (MCC$^{\mathbf{N}}$)** – diagonal of the residual covariance matrix: $\delta_i = \sqrt{\Omega_{ii}}$
- **Norma-scaled MCC (MCC$^{\mathbf{H}}$)** – *n-norm* of the Jacobian matrix rows:

$$\delta_i = \sqrt[n]{\left(\frac{\partial h_i}{\partial x_i} \right)^n + \ldots + \left(\frac{\partial h_i}{\partial x_n} \right)^n}$$

Hereafter, both ways of generalizing the MCC estimator are presented and subsequently tested in the selected case studies.

3.4.1 Normalized MCC

The MCC criterion is proposed to be normalized so that kernel size for each measurement is determined by the square-root of diagonal elements of the residual covariance matrix Ω. Therefore, the scaling factor for each measurement is calculated as:

$$\delta_i = \sqrt{\Omega_{ii}} \tag{3.52}$$

The same principle is used in LNRT only here it is the size of the kernels that is normalized, rather than the residuals. An analog method for the normalization of break-point of non-quadratic criterion was proposed in [9] and was later called the Schweppe-type GM-estimator. Although this scaling is not expected to entirely solve the problem of multiple GE, robustness of the common MCC criterion with a unique kernel bandwidth should be enhanced.

3.4.2 Norma-scaled MCC

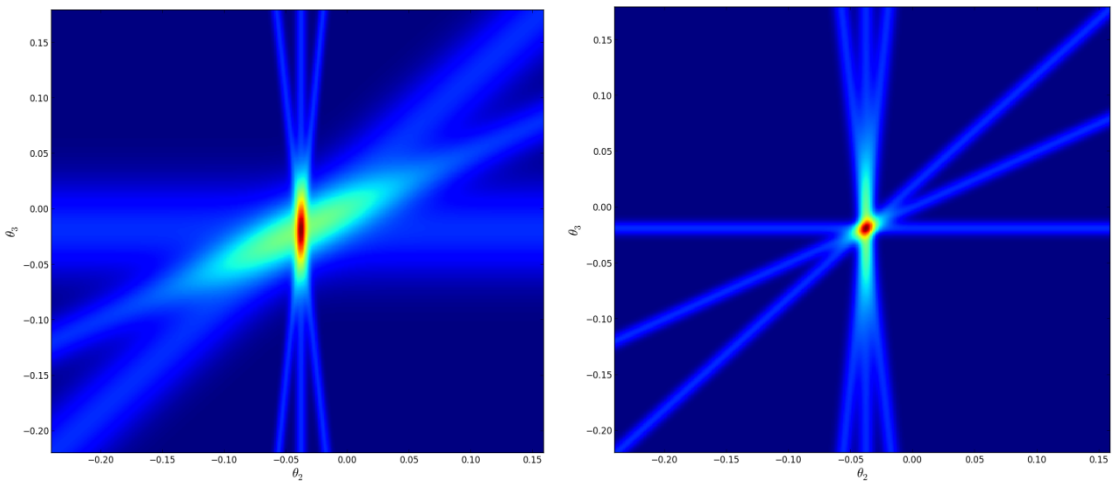
Here, a novel approach for kernel size scaling is proposed, motivated by equalization of contributions of each measurement to the cost function (in space of the state vector). More

specifically, Figure 3.24(a) shows Correntropy function for the 3-bus DC system (the same as in Figure 3.20(b)) which has impedance of line 1-2 10 times lower then of the other lines. Due to this, measurements P_1, P_2 and P_{1-2} create leverage points. The Correntropy function from Figure 3.20(b) assumes unique value of the kernel size ($\sigma_1 = \ldots = \sigma_m = 0.1$). It can be noted that although the same Parzen window bandwidth are considered, contributions to the Correntropy function of each measurement in the state vector space are obviously not equivalent. The reason for this comes from the Jacobian matrix numerical structure, defined by measurement type, network parameters and topology. Hence, when observing the linearized state estimation problem $z = Hx + r$, the Correntropy function between the measured and estimated values is calculated as:

$$J_{MCC}(x) = \sum_{i=1}^{m} e^{-\frac{(z_i - H_i x)^2}{2\sigma_i^2}} \tag{3.53}$$

where H_i is i^{th} row of the Jacobian matrix H. Therefore, contribution of each measurement to the Correntropy function is directly conditioned by the associated Jacobian matrix row. Accordingly, one proposes scaling of σ_i for each measurement according to the norm of the corresponding Jacobian matrix row:

$$\delta_i = \sqrt[n]{\left(\frac{\partial h_i}{\partial x_1}\right)^n + \ldots + \left(\frac{\partial h_i}{\partial x_n}\right)^n} \tag{3.54}$$



(a) MCC objective function before scaling.

(b) MCC objective function after scaling.

Figure 3.24: Scaling effect of Correntropy function for the 3-bus DC sample according to Jacobian matrix structure. It can be noted that after scaling, each kernel appeared with identical bandwidth in the state vector space.

In Figure 3.24(b), the effect of the proposed scaling of Parzen windows which uses (3.51) and (3.54) can be seen. If comparing Figure 3.24(b) with Figure 3.24(a), scaling obviously

decreases leveraging effect by equalizing contribution of each measurement in the Correntropy cost function. When compared to the normalization from previous section, this solution has a computational saving advantage since it does not require calculation of the covariance residual matrix.

In order to demonstrate benefits of the proposed generalized Correntropy criterion, the scenario of a single bad leverage point in measurement P_1, was considered. As already depicted in Figure 3.21, common MCC estimator (with unique σ) fails to identify bad data in the same scenario and ends in a local maximum. After applying the proposed vector (eq. (3.54)) for Parzen windows scaling, masking effect is eliminated and Newton's method manages to find the global maximum, as obvious from Figure 3.25. Also, comparing to Figure 3.21, it is significant to observe equilibration of particular measurements contributions in the objective function. This result is very significant since indicates that the proposed scaling makes the MCC estimator robust to most demanding scenarios of single gross errors. In fact, it eliminates probability of failure in case of a single bad data. In the following section, behavior of the generalized MCC estimator is tested on multiple GE scenarios.

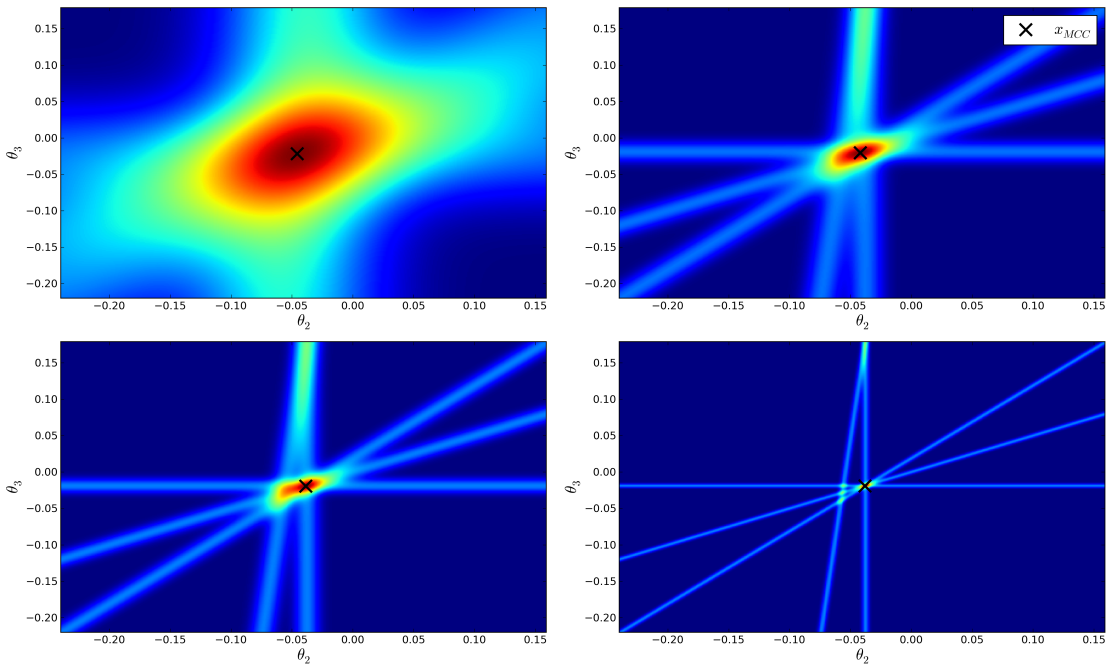


Figure 3.25: 2D projection plot of the generalized Correntropy function for the 3-bus DC system with bad leverage point in measurement P_1. It is of key importance to note benefits of the Parzen window scaling in comparison to Figure 3.21.

3.4.3 A Comparative Study

3.4.3.1 Case Study - the 4-bus System

Both proposed generalized MCC estimators are expected to improve robust performances of a general Correntropy criterion. In order to verify this, a comparative study is performed on the 4-bus system case study from section 3.3.2. The summarizing results regarding the efficiency in bad data identification are given in Table 3.13. Both MCC^{N} and MCC^{H} significantly outperform the common MCC. When compared to LNRT, almost the same efficiency is achieved.

Table 3.13: Efficiency of identification of 2 GE with LNRT, MCC, MCC^{N} and MCC^{H} considering 10000 measurement scenarios of the 4-bus system. Although the generalization of the MCC has contributed to robustness, the LNRT method still remains most efficient.

	Wrong	**Correct**	**Efficiency**
LNRT	166	9834	98.34%
MCC	300	9700	97.00%
MCC^{N}	174	9826	98.26%
MCC^{H}	179	9821	98.21%

3.4.3.2 Case Study - 24-bus System

The state estimation methods are tested again for 3000 measurements scenarios of the 24-bus system from Section 3.3.3. The results from Table 3.14 once again reveal significant benefits gained by proper scaling of the Parzen windows. Moreover, in this case, the generalized MCC estimators MCC^{N} and MCC^{H} both slightly outperform the conventional LNRT method in bad data identification. According to all case studies performed, one can conclude the following:

Table 3.14: Efficiency of identification of 2 GE in active power measurements for LNRT, MCC, MCC^{N} and MCC^{H} in 3000 measurement scenarios of a 24-bus system.

	Wrong	**Correct**	**Efficiency**
LNRT	190	2810	93.67%
MCC	289	2711	90.37%
MCC^{N}	159	2841	94.70 %
MCC^{H}	155	2845	94.83 %

- Proper scaling of the Parzen windows brings unquestionable improvement in robustness of the MCC estimator, when compared to the basic Correntropy criterion which considers identical values for all Parzen windows $\sigma_i, \forall i = \{1, \ldots, m\}$.

- Generalized MCC estimators MCC^N and MCC^H reach robustness that is approximately on the same level as the LNRT.
- The MCC^N and MCC^H are both generalized MCC estimators that use different scaling methods. Provided efficiency of the estimators is almost the same. The only advantage of MCC^H over the MCC^N estimator is in fewer calculation requirements in regard to the covariance residual matrix Ω.
- Generalization of the MCC estimator contributes to its robustness. However, failures are still probable to occur, due to Newton's algorithm + kernel annealing solver, which remains able to be miss-led to a local maximum of the Correntropy function, instead of the global.

3.5 Conclusions - MCC State Estimator

One of key functions of a bad data processor within a state estimator is to detect the presence of erroneous measurements, identify and finally eliminate them from the dataset or (if possible) correct them. This problem is approached in this thesis from an information theoretic point of view with the main purpose of extracting the maximum information possible from the available measurement dataset. A criterion implementing this idea is the MCC – Maximum Correntropy Criterion – a function defined in the Information Theoretic Learning context as a robust measure of similarity. Therefore, Correntropy is used here in order to measure the similarity between the distribution of the measured values (provided by the SCADA) and the estimate of the same measured values.

Another way of providing an interpretation of the process is to understand that Correntropy is associated with a metric (CIM); therefore, one can say that the Euclidean distance (simple, as in LS, or normalized, as in WLS) is to be substituted by the CIM distance (minimizing CIM is equivalent to maximizing Correntropy). However, this interpretation is insufficient: it obscures the true change in paradigm brought up by the new approach, which is related to understanding the SE problem as a problem of information, patterns and outliers.

Unlike the WLS state estimator, which considers only a L2-type norm and, in fact, minimizes the variance of the weighted error distribution, the proposed Maximum Correntropy Criterion (MCC) state estimator acts as if changes in the metric norms occur, from L2 through L1 to L0 as the errors increase. Due to advantageous properties of Correntropy and CIM distance, by generalizing the correlation concept and not depending on an Euclidean distance, the novel MCC state estimator inherits these benefits and contributes with substantial enhancement in robustness to bad data. One can say that the MCC estimator behaves equivalently to the WLS when the errors are small, relative to the kernel size, i.e. when the errors remain under the quadratic part of the Gaussian function. On the other hand, if a measurement error resides far from the origin and lies in the region governed by the L0 norm, the Correntropy becomes simply insensitive to its value, i.e. it is treated as an outlier. In this case, this error is in practice identified as an element of bad data and its influence is naturally eliminated from the estimation process. This introduces a new

degree of simplicity in the estimation process, because bad data no longer need to be eliminated from the data set and, as a consequence, the matrices are no longer needed to be recomposed – the bad a data remain but their influence becomes negligible - the most desirable consequence.

In the preliminary tests, the Correntropy was maximized with an evolutionary particle swarm optimization (EPSO), which has availed extreme robust features of the novel objective function in any GE scenario, and thus outperformed all the conventional robust estimators, which are prone to identification failure. The MCC estimator did not reveal any cases of failure even if the strongest leverage points were corrupted. In general, one has also realized that with strong robust features, the novel criterion works well also with a very demanding objective function, due to multiple local optima that emerge in the presence of bad data. In any case, the powerful EPSO method did not have any difficulties in finding the global maximum which always corresponds to the true solution, ignoring the deteriorating effect of GE (unless GE corrupt critical k-tuple). Unfortunately, the EPSO cannot be considered as an industrially-acceptable solution procedure: it does not allow scalability to large systems, due to the computational burden that emerges with the growing size of the problem.

As an alternative solver, a computationally more acceptable and suitable for the real time environment, the Newton-Raphson method is proposed. However, one needs to be aware of the limitations of the Newton's method, which do not allow it to be applied on a multi-modal objective function which is non-concave/convex on all its domain. In order to address this issue, taking advantage of the free parameter σ which represents the width of the Parzen window, has been suggested. By tuning the kernel size and carefully annealing it, the objective function landscape changes from a well behaved concave quadratic shape to a topology with eventually numerous local optima. With these specific properties in mind, a procedure of smooth kernel annealing that enables fast and reliable convergence can be proposed.

In other words, a sufficiently large σ reduces the MCC criterion to the WLS, i.e. it makes the optimization easy and fast-converging (solving the traditional problem, actually). Subsequently, the Parzen window size is smoothly annealed thus allowing a fast and reliable convergence of the Newton's solution. **It is very important to emphasize that the optimization of the Correntropy function embodies in itself all bad data processing functions: detection, identification and elimination, without any additional calculations or re-processing.** In addition, it is important to understand a novel detection mechanism enabled by the MCC estimator. If the solution does not change after a few iterations with an initial large kernel size successively annealed, one can deduce that the available measurement vector is healthy, i.e. there is no GE corruption. In that case, the MCC will produce a result equivalent to the WLS and the solution initially obtained is final. On the other hand, in the presence of gross errors, the solution deviates from the initial one. Then, a distilling process starts, the Parzen windows are decreased in multiple steps and all erroneous measurements are progressively eliminated.

Exhaustive tests on various case studies revealed that the MCC state estimator may fail in some GE scenarios and end up converging to a local maximum, far from the true state vector. The iterative solution, for example, may be misled and get trapped within some concave region during

the annealing process which finally ends up as a local optimum. Statistically, these scenarios are not very probable to occur, but it is interesting to observe that they coincide with the failure cases in most of the conventional methods, i.e. they are always conditioned with multiple interacting and conforming GE, or even more if one of them is a bad leverage point. Therefore, one concludes that MCC robust performances (using Newton as the optimizer) are very similar to estimators that optimize a non-quadratic function of the measurement residuals as well as the well-known method of the largest residuals elimination.

As a final conclusion, one can say that the proposed Correntropy objective function possesses extreme robust features that place it beyond the known robust estimation criteria. However, the adopted optimization procedure - Newton's method + Parzen window adaptive annealing, is unable to extract the information maximum from every case, i.e. identify correctly every bad data. The initiation with the large kernel size equates MCC to WLS and by doing this it falls again in the well-known trap of the masking effect among residuals the multiple interacting GE may cause. In order to eliminate this effect, generalized Correntropy criterion is proposed which defines ratios between each kernel size, instead of the unique kernel size for all the measurements. These ratios can be calculated from the Jacobian matrix structure or according to diagonal of the residual covariance matrix. In both solutions some robustness is gained, especially in the cases of single bad leverage points, when probability of failure is reduced to zero. However, in cases of multiple GE, generalized MCC keeps similar robustness level as the LNRT, so it is still prone to failure in incident and conforming GE scenarios.

Although one can say that the MCC state estimator is just another non-quadratic criterion and does not provide any significant robustness enhancement, there are multiple contributions of key importance which can be summarized as the following:

- ◇ The Correntropy function is infinitely and continuously differentiable on all of its domain, unlike other non-quadratic criteria proposed in the literature.

- ◇ Global maximum of the Correntropy function always corresponds to the true solution

- ◇ Most of the non-quadratic criteria lead to formulations classified as Mixed Integer Non-Linear Problems (MINLP) due to the required binary unknowns. The MCC estimator does not require any binary variables, which classifies it as a normal non-linear problem (NLP).

- ◇ The MCC estimator offers a strong theoretical justification for all the previous attempts to modify the quadratic objective function focused on gaining robustness in the estimation process. The perspective of SE gained from the information theory point of view takes the (hidden) true values as a reference to measure errors, while the classical methods work with residuals, which take a set of estimated values contaminated by errors as references. The fact that global maximum of the Correntropy function always corresponds to the true solution, confirms this property.

- ◇ The MCC estimator opens a new paradigm in state estimation: instead of looking at a snapshot of measurements (with errors) and combining it with Kirchhoff Laws, it looks at the measurement set in terms of combination of a multidimensional signal with a small noise (small errors) and outliers (the GE) – and the process aims to reconstitute the original signal by identifying and correcting the GE.

In summary, the adoption of Correntropy function in SE brought many novelties and is considered to be a strong advancement, based on a theoretical principle. However, the computational procedure using meta-heuristics is complex and unfeasible to be used in real time – therefore, the Newton procedure was proposed. Although not fully robust, it nevertheless opened some interesting, new avenues. In the following chapter, we will proceed to develop a complementary approach that will compensate the limitations introduced by the proposed Newton method.

Chapter 4

Auto-associative Measurements Screening and the Hybrid MCC Estimator

The former section proposed a novel state estimator, based on the MCC, which establishes a new reference for the estimation process founded on information theoretic paradigms. However, the maximization procedure (based on the Newton method and the Kernel annealing strategy) has proven to be insufficient, by itself alone, in extraction of all possible benefits from the novel criterion. Put simply, the procedure is fast and accurate, but not sufficiently reliable in all the GE scenarios.

The main reason for this may be found in the optimization process when using the Newton method. This method depends on definition of a good starting point for the iterations. It is known that the Newton iterations for maximization can only assure convergence if they develop inside a concave region of the objective function landscape and if the global optimum is the maximum in that region. The smooth annealing process cannot assure this at all times. The annealing process has been demonstrated to induce a continuous change in the shape of the objective function and the iteration step could remain trapped inside a wrong concave region from which it cannot escape. Although this may be a rare phenomenon, its occurrence nevertheless undermines the confidence one might have in the MCC method.

Due to these circumstances, the envisaged solution is to provide a method that would capture the manifold from the available measurements data set and retain a stable reference point that would enable a much more robust and efficient state estimation.

Therefore, an unconventional solution that intends to escape the classical perception of the problem as mathematical regression in the measurement space is introduced in this section. A special kind of Artificial Neural Networks (ANN), called Auto-associative neural networks (AANN), are envisaged as a tool to extract all the relevant information from the available dataset, learn the manifold and store it in their weights. These trained neural networks are intended to be used in real time processing of available data with the main purpose of performing a data screening

process. If they are able to perform as expected, the AANNs can be used as a strong reference point of the MCC estimator in a hybrid state estimator. Thus high robust features are expected to be gained, without the engagement of some excessive computational efforts that threaten the scalability of the method.

4.1 Building Blocks - Auto-associative Neural Networks

4.1.1 Autoencoders

Auto-Associative Neural Networks (**AANN**), also known as ***Autoencoders***, are feedforward networks trained to mirror the input space S in their output. These neural networks are proposed in order to achieve an approximation of one-to-one mapping between points of a space of dimension m and a space of dimension n (where $n < m$ without loss of generality). Specifically, in an autoencoder, the first half of the neural network approximates the function f that maps the input space to the space of compressed encoding S', while the second half approximates the inverse function f^{-1}. In other words, an autoencoder is trained to display an output equal to its input. During the training process it stores the information about the data manifold in its weights.

Figure 4.1 depicts a simple neural network with a smaller single middle layer. In general, there is no special requirement for the architecture of an AANN, although the two halves of the network should be built symmetrically. However, the most widely adopted architecture contains a single hidden layer due to the fact that networks with more hidden layers have been proved to be difficult to train [134]. In general, the accuracy of the ANN model is achieved at the cost of training demand.

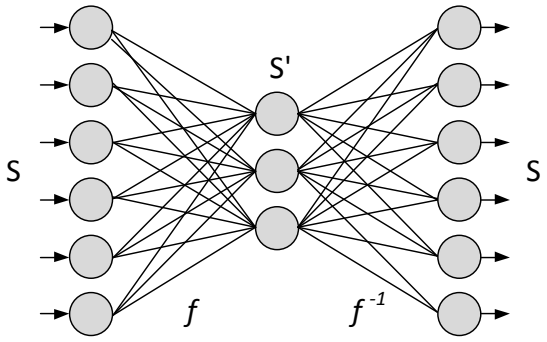


Figure 4.1: Architecture of a 6-3-6 autoencoder with a bottleneck inner layer and input and output layers of the same dimension. The inner layer contains a compressed set of values that encode values in S in a reduced dimension space S'.

Moreover, there is no any way to define an adequate compression rate (measured as the ratio between the number of neurons in the smallest middle layer and the number of neurons in the inputs/output layers) *a priori*. In addition, there is an obvious trade-off between the compression

rate and the accuracy of data reconstructed through the second half of an autoencoder. The decision about the space compression rate will evidently depend on the nature of the problem and it is made in the present-day practice simply by trials and errors.

One of first applications the autoencoders were proposed for, was the image compression. Specifically, the first half of AANN maps the input to a reduced dimension space and the signals available in the middle layer are used to store information about the image. In addition, weights of the second half are required in order to reconstruct the originals [135], [136].

Furthermore, the application is extended to distinct processing techniques such as identification and pattern recognition [137]. For instance, face images could be identified and clustered according to sex, distinguished from non-faces, etc. [138]. A different application has been presented in [139], [140], for the reconstruction of missing sensor signals. This belongs to the type of missing data problems, where autoencoder properties have been used to reconstruct some missing input data in such a way that the reconstruction appears reproduced in the output, by minimizing the function of the input-output error.

It is interesting to observe that an autoencoder with one hidden layer and linear activation functions performs the same basic information compression from space S to space S' as ***Principal Component Analysis*** (**PCA**) [141]. A substitution of linear with non-linear activation functions, upgrades the autoencoder to the so-called non-linear PCA [142]. With non-linear activation functions and multiple layers, the autoencoders chart the input space on a non-linear manifold so that an approximate reconstruction is possible with fewer errors [143]. Moreover, the PCA does not easily show how to do the inverse reconstruction, which is straightforward with autoencoders.

4.1.2 Missing Data Reconstruction and Bad Data Filtration with Autoencoders

The key important capability of autoencoders, intended to be adopted in this thesis for the state estimation purposes, is the reconstruction of missing data. This ability was initially recognized in [139], in the context of reconstruction of missing sensor signals. The main idea is to use the autoencoder properties for the reconstruction of missing input data so that the reconstruction appears reproduced in the output, by minimizing the function of the input-output error. In other words, an optimization method is used to find the missing input that is most coherent with the output provided by the autoencoder.
The training of an autoencoder is envisaged to be performed offline with an appropriate dataset. Afterwards, if a new vector of the same data type is presented to the trained autoencoders, the following two scenarios can occur:

$$\begin{aligned} &\textit{Vector is consistent with the learned cluster} && \Longrightarrow && \textit{small input-output error} \\ &\textit{Vector is an outlier regarding the learned cluster} && \Longrightarrow && \textit{large input-output error} \end{aligned}$$

According to this behavior, the following can be concluded: if an incomplete pattern is shown to the autoencoder, (missing components are replaced by random values), a large error will appear between the input and the output vectors. Moreover, a trained autoencoder is expected to associate the erroneous or missing topology with the true one, i.e. to project the available vector to the

learned manifold, as depicted in Figure 4.2. If successful, this methodology could equally be used for both the missing data reconstruction as well as the correction of bad data. In any case, a reconstruction task is set to search for the missing value that will minimize a function of errors between the input and the output vectors.

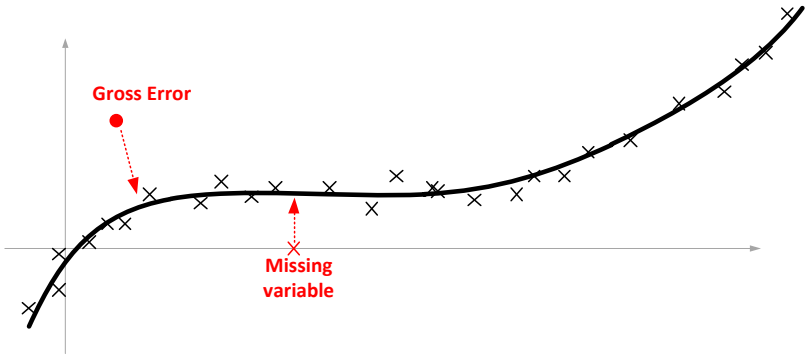
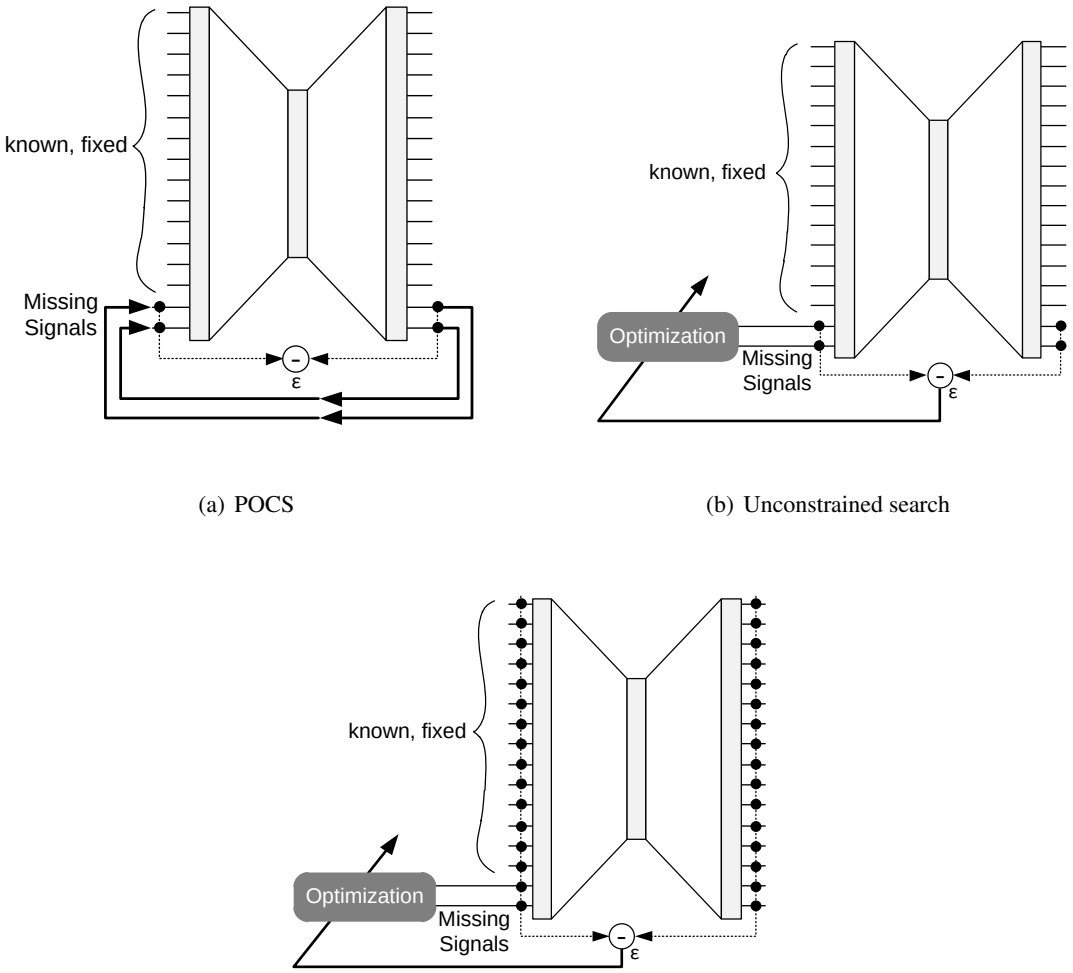


Figure 4.2: Illustration of autoencoder abilities for correction of gross errors or reconstruction of missing signals.

Depending on the method used for the search of the missing value, three basic approaches are distinguished:

- ***POCS – Projection Onto Convex Sets*** [139]: This model uses alternating linear projections on the input and output space to converge to the assumed missing value. If an input signal is missing, its variable can be set to a random value and this will produce a mismatch between the input and the output. Iteratively reintroducing the output value in the input will converge to a value that minimizes the input-output error (Figure 4.3(a)).
- ***Unconstrained search***: this model requires an optimization algorithm to minimize the input-output error. The unconstrained search controls the convergence by the error in the missing signals (Figure 4.3(b))
- ***Constrained search***: unlike the unconstrained search, the constrained search controls the error in all the outputs of the autoencoder (Figure 4.3(c)). For instance, in [144], a genetic algorithm is used to optimize the error function.

The same approach of missing signal restoration could be extended to the problem of correction of bad signals. However, this task is still substantially more demanding since the inputs which are to be reconstructed are not known *a priori*, as in the case of a missing signal. First, the bad inputs need to be detected and then the erroneous inputs need to be identified. Therefore, before the process of input reconstruction begins, the autoencoder needs to detect and



(a) POCS

(b) Unconstrained search

(c) Constrained search

Figure 4.3: Illustration of the algorithms. (a) *POCS* - the missing input is iteratively fed from the corresponding output until convergence is achieved; (b) *Unconstrained search* - an optimization algorithm searches for input values that minimize the input-output error of the missing signals; (c) *Constrained search* - An optimization algorithm searches for input values that minimize the input-output error of all signals.

identify the erroneous variables, which requires significant training accuracy and robustness. In that case, one-to-one correspondence is needed between the input-output error of the autoencoder and the true gross errors. After the restoration process is done, the error can be quantified and the erroneous input substituted with the estimative provided by the autoencoder. In Figure 4.2, the schematic illustration shows how gross errors alienate the vector from the expected manifold and the autoencoder is called to project this outlier to the closest point on the manifold.

In order to improve the autoencoder's abilities for projection of outliers onto the manifold, Vincent *et al.* in [145] introduced a new training principle which established the so-called *Stacked*

Denoising Autoencoders. The adjective *stacked* comes from a principle of unsupervised training which is performed layer-by-layer. Each layer is first trained to produce a representation of the observed patterns for the subsequent layer, based on the representation that it receives as an input from the layer before and by optimization of a local unsupervised criterion. This training provides initialization of weights layer-by-layer for the global training of an autoencoder with deep architecture. Moreover, the adjective *denoising* addresses the idea of autoencoder training for the purposes of reconstruction of a repaired input from an artificially corrupted one. More specifically, the input data are first submitted to corruption process which randomly corrupts information about some of the components. The autoencoder is than trained to reconstruct erroneous inputs. The training of this type produces a very robust, stable and accurate autoencoder, which is used in [145] for image classification experiments. This approach is adopted and additionally improved in [146] and [147].

4.1.3 Autoencoders in Power Systems - an Overview

Until now, there were not that many applications of auto-associative neural networks in power systems. However, there are some publications which address the issue of missing signal restoration. In [148], some models for missing sensor signal restoration are discussed in the context of non-linear power plant operation. In addition, in [149] the autoencoders are coupled with a wide area state predictor, as a part of a wide area controller, in order to restore the missing signals. Report of enviable performances of autoencoders in diagnosing faults in power transformers has recently been published [150].

In regard to the power system state estimation, there are some very interesting applications of ANN, quite relevant to the work that is reported in this thesis. Salehfar *et al.* have proposed an ANN with input data patterns identical to the target (output) patterns, and used it as a pre-estimation filter for bad-data detection and identification in PSSE [151]. Moreover, in order to maintain observability, identified bad data are not rejected but replaced with estimates provided by ANN.

A similar approach was reported in [152], where a two-stage ANN was proposed. The first part is trained to estimate the system state from the available measurements, while the second part takes over the system state vector and reconstructs the measurement vector. All of this is envisaged for the purpose of bad data detection and identification. Although these methods seems to work in tested cases of multiple bad data, the authors report a need for accuracy enhancement as well as high dependence of ANN performances in architecture and training settings.
The third and the most recent work about ANN as a filter for measurements prior to the state estimation process is reported in [153]. The proposed concept uses historical measurements as an input for forecasting the subsequent measurement vector. In cases when significant deflection is detected, bad data are identified and corrected by the same ANN.

The main motivation for adopting autoencoders for power system state estimation purposes is to maintain the information-theoretic approach and, in the meantime, escape the classical perception of the problem as mathematical regression in the measurement space. The training

is to be performed offline, with the available historical measurements dataset. Afterwards, a trained autoencoder is envisaged to be used in real time in order to associate the erroneous or missing electric signals with the true ones or at least indicate the outlier signals. Also, if some measurements are missing from the acquired set, the autoencoders could reconstruct them according to available measurements. If it manages to meet these expectations, the autoencoder could be used as an auxiliary tool in power system state estimation, a pre-filer which could provide a state estimator with filtered measurements complete and free of GE.
In addition, the autoencoders have been found to be very suitable for learning the manifold implicit in the link between the measurements and the topology states. In that case, they are expected to process two types of signals that exist in power system in parallel: analogue (power, voltage or current measurements) or digital (breaker statuses). Chapter 5 of this thesis discusses this issue in particular and focuses more on topology estimation problem.

The proof of concept developed in this thesis about the use of autoencoders for the reconstruction of missing signals in power systems was published in [154].

4.2 Reconstruction of Analogue Measurements with Autoencoders

This section presents the capabilities of autoencoders to perform data screening in the two following ways:

- ▷ Restoration of missing electric measurements,
- ▷ Detection, identification and correction of bad measurements.

During the training phase, which is done off-line, the autoencoders learn all the information hidden within the available measurement database. This information is kept stored in autoencoder weights and is being used later in real time in order to filter out the potentially corrupted measurements received from the SCADA. The possible ways of corruption are gross errors, or simply the absence of measurement signal. This problem is unified under terms like measurements restoration, reconstruction or screening. However, there is a key difference between the two issues: missing measurements or erroneous measurements. The missing values are by definition already detected and identified. What remains to be done is to reconstruct the value in accordance with the other available measurements. Alternatively, in case of gross error, the measurement first needs to be detected, identified and then finally corrected. Therefore, the bad data case is found to be more demanding and will be considered in this thesis separately and in more detail.

The capabilities of autoencoders for processing and reconstruction of electric measurements are tested on the 4-bus system case study from Chapter 4 (described in detail in Appendix A.1). The autoencoders architecture is defined by the reduction of the hidden layer as cca. 60% of the input size. Neurons within the hidden layer use activation function, while the output neurons are linear.

4.2.1 Generation of Pseudo-measurements with Autoencoders

Pseudo-measurements in a power system are generated according to available information about the system in order to improve the redundancy or restore observability in certain parts of the system. They are used together with all other measurements. However, unlike the common ones provided from the SCADA, these are obtained artificially prior to the estimation process. Hence, pseudo-measurements are generated in order to submit missing or extra information to the state estimator and improve the final result. Their calculation derives from external processes such as historical data, prediction procedures, load curve assessment, calculations on an external network, etc.

The importance of generation of external pseudo-measurements was discussed, for instance, in [155]. Also, in [156], the generation of pseudo-measurements is based on forecasted load data in systems with distributed generation. In [157] pseudo-measurements in distribution systems were generated using probabilistic models on historical load data. The generation of pseudo-measurements as an internal process, however, is not usual and has not been considered.

The idea proposed in this thesis is to train AANN to learn the manifold of available measurement dataset. That manifold is generally defined by various factors: Kirchoff laws, load distribution, system parameters, topology, etc. All this information resides hidden in the historical dataset, and the task of the autoencoder is to extract it. After off-line training, the information about the manifold is stored in autoencoder's weights. In real time, the measurement vector acquired from the SCADA is presented to the trained autoencoder. If the measurement vector is consistent with the learned data, the input-output error will be insignificant. On the other hand, a large error will indicate that the input vector is an outlier due to the learned manifold. These properties are to be used in order to re-project this outlier vector back into the cluster and thereby correct the corrupted inputs.

In section 4.1.2, three basic concepts for the reconstruction of missing signals with autoencoders have been presented: POCS, the constrained and the unconstrained search. In order to adopt the most suitable concept for a particular purpose, all three were tested on a set of sampled cases for the 4-bus system. In total, 20000 scenarios were generated for active and reactive power measurements (8 active power and 8 reactive power measurements), where 10000 were used for training and the other 10000 for testing. In test database, one of the inputs per each scenario is randomly specified as missing and initialized at random value. For constrained and unconstrained searches a meta-heuristic optimization method EPSO has been adopted for the optimization of objective function of input-output errors. Hence, POCS, EPSO-unconstrained and EPSO-constrained were tested on restoration of missing measurements of active and reactive power. The preliminary objective function is defined as a sum of squares of input-output errors (MSE):

$$\varphi_{MSE}(\varepsilon) = \frac{1}{N}\sum_{i=1}^{N}(z_i - \hat{z}_i^{out})^2 \tag{4.1}$$

where:

$\hat{z}^{out}$ estimated output vector of the neural network, as a function of estimated missing input vector $w(\hat{z}^{in})$ (Figure 4.10),

N number of all the inputs in case of constrained search or number of missing inputs in case of unconstrained search.

The PDF of input-output errors in 1000 test scenarios is shown for all three methods in Figure 4.4. The POCS algorithm is straightforward and simple, but in the case of multiple missing values it becomes insufficiently accurate and eventually does not converge. The unconstrained EPSO search provides results which are, due to accuracy, quite similar to POCS. The constrained search significantly outperforms both the unconstrained search and POCS. Moreover, it is interesting to observe that input-output errors of the constrained search are even lower than errors obtained by inputs free of GE. In other words, the constrained EPSO search is able to project the input vector even closer to manifold than the autoencoder itself. Moreover, the constrained search does not require any additional computational effort when compared to unconstrained search and provides significantly more valuable results. Therefore, it is straightforward to choose the approach of constrained search (with EPSO algorithm) for the purposes of restoration of missing or bad measurements.

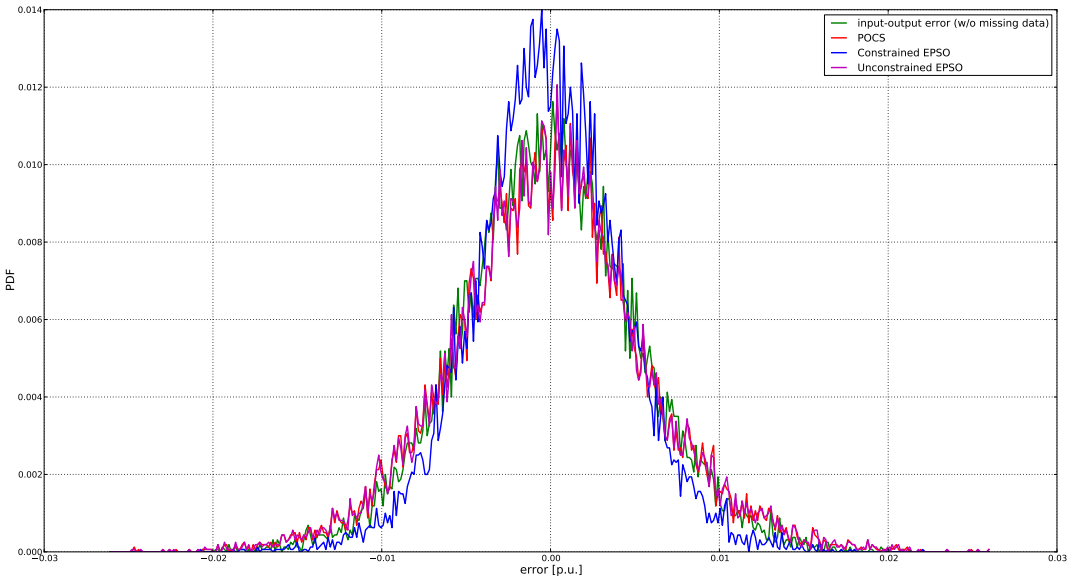


Figure 4.4: PDF of input-output errors for missing input restoration with: POCS, constrained-EPSO search and the unconstrained-EPSO search. The constrained-EPSO search is shown as the most accurate, also in comparison with the true inputs, free of GE.

4.2.2 Gross Errors Detection, Identification and Correction with Autoencoders

In the former section, it has been shown that the autoencoders are able to capture information between the electric power measurements. In this section, the information stored within autoencoder weights is intended to be used to detect, identify and correct bad data. The problem

of bad data analysis is addressed in more detail in Chapter 4. In addition, a novel MCC estimator is proposed that sheds some new light on the state estimation problem, based on information theoretic paradigms. However, the Newton's method with kernel annealing strategy does not allow the global optimum in all the scenarios to be found, i.e. it fails to identify some of bad data.
The MCC estimator is designed to maximize the information content between measured and estimated values. This section intends to test the abilities of autoencoders to estimate the measurements in order to equip the MCC estimator with more robust features.

The preliminary tests of auto-associative neural networks (autoencoders) in the former section hinted their ability for reconstruction of missing measurements. Moreover, they could be also used to filter-out gross errors, only if they are able to identify them. More specifically, if the problem is observed in a more general light (i.e. all the GE are identified), each corresponding erroneous measurement may be proclaimed as missing and afterwards corrected through the reconstruction process.

The following analysis investigates first capabilities of autoencoders, at least as GE detectors. In other words, if the input measurement vector is an outlier from the learned manifold, a trained autoencoder is expected to result in a large input-output error. If additional outliers in the input-output errors correspond directly to bad data, this property would satisfy the identification and, to a certain extent, the quantification of GE.
The 4 bus case study is considered with 20000 scenarios of active and reactive power measurements. The autoencoder is composed of a single hidden layer which is reduced to cca. 60% of the input vector size, so it results with an architecture of 16-10-16.

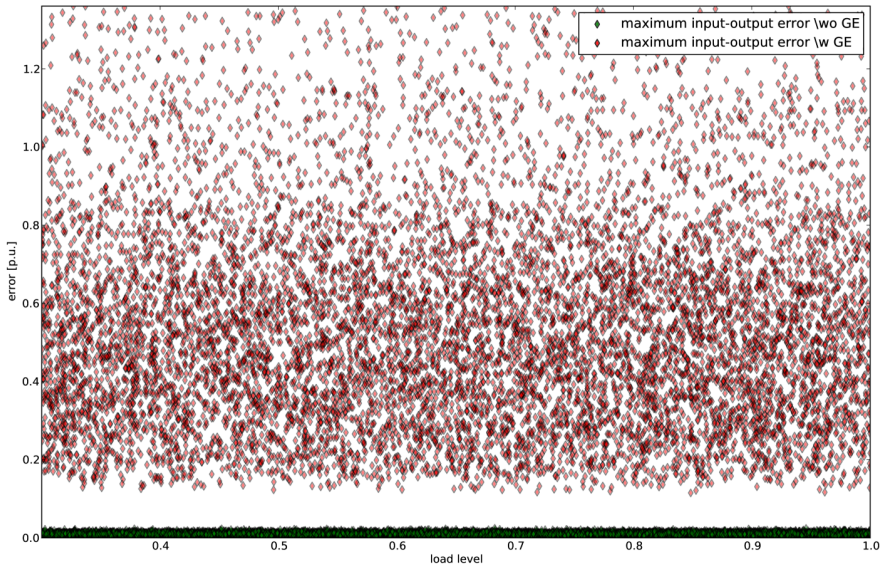


Figure 4.5: Scatter plot of maximum absolute input-output errors of the autoencoder (depending on load level) for test cases with and without the gross errors. It is significant to note that the cases infected with GE are easy distinguishable, i.e. the autoencoders can be used as GE detectors.

The autoencoder was trained with 10000 samples of measurement vectors which contain normally

distributed errors with a variance of 0.005 p.u. Two testing datasets were considered: one contains the same Gaussian errors as the training set and the other includes an additional 1 GE per each scenario. Gross errors are generated with a uniform distribution in the range 20 to 100 standard deviations of the normal distribution. Figure 4.5 shows a scatter plot of maximum absolute input-output errors depending on the system load, for both testing datasets, with and without GE. It is significant to note a high increase of maximum error in the presence of gross errors. Also, there is no overlap between maximum error with and without GE, so detection is in this case able to be performed with a 100% efficiency. Regarding the system load, one needs to say that errors do not show any dependence on load level in the system.

Now, it is interesting to observe errors of each particular input in order to examine the potential correspondence between GE and input-output errors of the autoencoder. If such correspondence is confirmed, it would mean that the autoencoder is able to identify and correct GE by projecting it directly onto the manifold. Therefore, the errors corresponding to inputs contaminated with GE and errors for inputs that are free of GE can be separated. Thus input-output errors are now plotted for each input in Figure 4.6. It can be observed that a single GE causes a significant smearing effect, i.e. distorts most of other outputs which correspond to true input measurements. Therefore, if attempting to identify and filter out the bad data, such autoencoder behavior would be far from acceptable. Put simply, an ideal scatter plot (compared to the one from Figure 4.6) would contain blue spots as low as possible, while the red colored errors should be distinguishably larger in range of gross errors. This type of autoencoder would be able to perfectly project every outlier onto the manifold and to also identify and correctly quantify each GE in the process. Obviously, the autoencoder created here does not have such identification properties. It is only able to perform gross errors detection, while some enhancements need to be adopted for the identification and correction of GE.

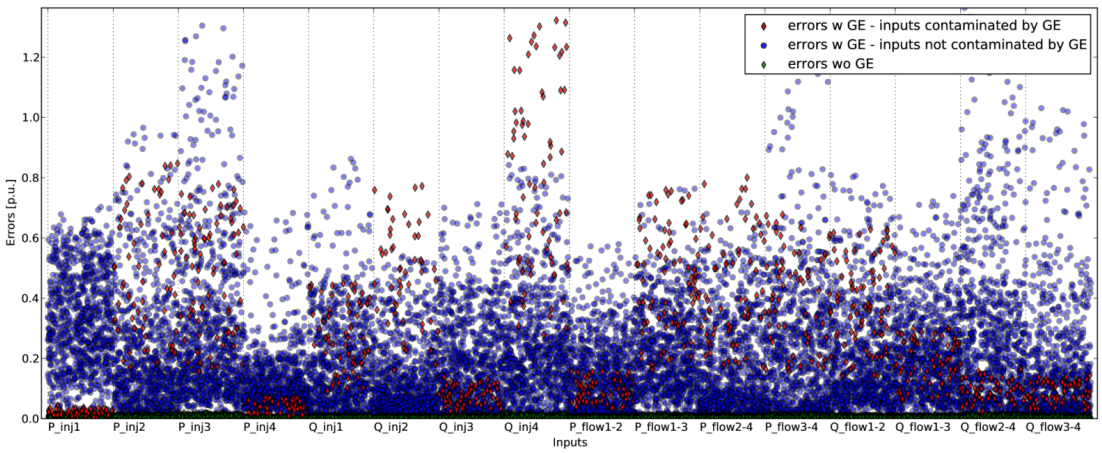


Figure 4.6: Scatter plot of absolute input-output errors per each input with and without gross errors. It is significant to note that input-output error of the particular input does not correspond to presence of gross errors in that input. An ideal autoencoder should keep blue errors significantly lower than the red ones, and the red ones should be higher, representing the gross errors.

4.2.3 Gross Errors Detection, Identification and Correction with Denoising Autoencoders

The autoencoder training uses back-propagation algorithm to minimize the sum of squared residuals between the output and target vectors, where target vectors are identical to the input ones. This methods is obviously satisfactory for the purposes of learning the data manifold, but when an outlier vector is presented in the input, it results in an outlier also in the output, i.e. it does not manage to project the input back onto the manifold. In order to equip the autoencoder with such skills, the previously presented concept of ***Denoising Auto-Associative Neural Networks*** (**DAANN**) (or simply denoising autoencoders) has been adopted. The idea is to intentionally corrupt the training input vectors with GE, while the target dataset is kept as original, free of GE. The principle of the method is illustrated in Figure 4.7. This way, the autoencoder is expected to learn not only the manifold, but also how to re-project an outlier back to the original data manifold. In other words, DAANN is taught to be able to reconstruct the original measurements vector from the vector contaminated by GE. This concept was proposed recently in [146] and [147] for the purpose of improving denoising in image processing. In this thesis, it was adopted for the first time for screening the input measurements for power system state estimation.

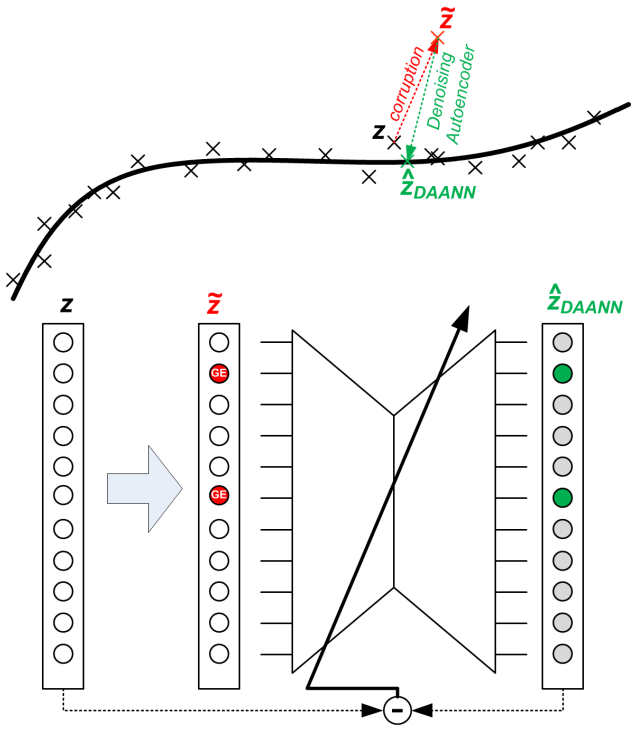


Figure 4.7: Training principle of a Denoising Autoencoder. Available measurements vectors z are concentrated around the manifold. A corrupted measurement vector $\tilde{z}$ is generated by introducing GE. The denoising autoencoder is thought to re-project corrupted measurements back onto the manifold and provide estimated vector $\hat{z}_{DAANN}$.

A training set for denoising autoencoder contains 10000 scenarios, where each input is

corrupted with 2 randomly generated gross errors, while the target set contains the same measurements but free of gross errors. Since all the input vectors contain GE, this training set has a corruption level of 100%.

A scatter plot of errors per each input for the autoencoder trained for denoising is shown in Figure 4.8. A significant difference can be noted in comparison to Figure 4.6. Specifically, by introducing the denoising concept, the desired performance of projection is significantly improved. Errors of non-corrupted inputs (blue spots) are significantly decreased, while errors of the corrupted inputs are kept on the level of true gross errors. In other words, the denoising autoencoder learned not only the data manifold, but also how to project the outlier inputs back onto the manifold. Unlike the normal autoencoders, such scatter-plot is already much closer to the ideal one and it also looks promising for use in identification of bad data.
Therefore, the abilities of a denoising autoencoder to identify 2 GE due to a predefined threshold were tested on the same database. It managed to correctly identify both GE in 95.6% out of 10000 cases. However, 15.2% of these suspected sets contained also some misidentified true measurements. On the other hand, the normal autoencoder has proven to be completely unusable for that purpose with a poor identification efficiency.

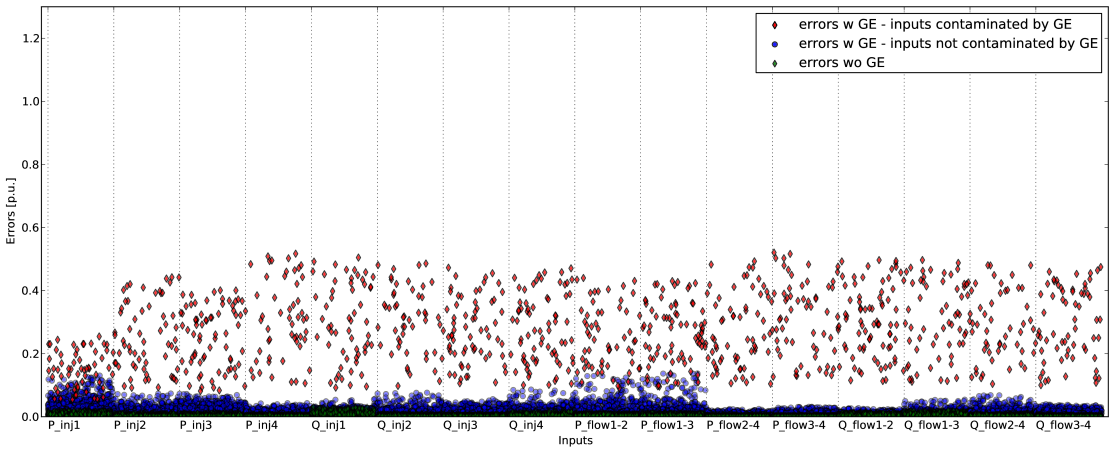
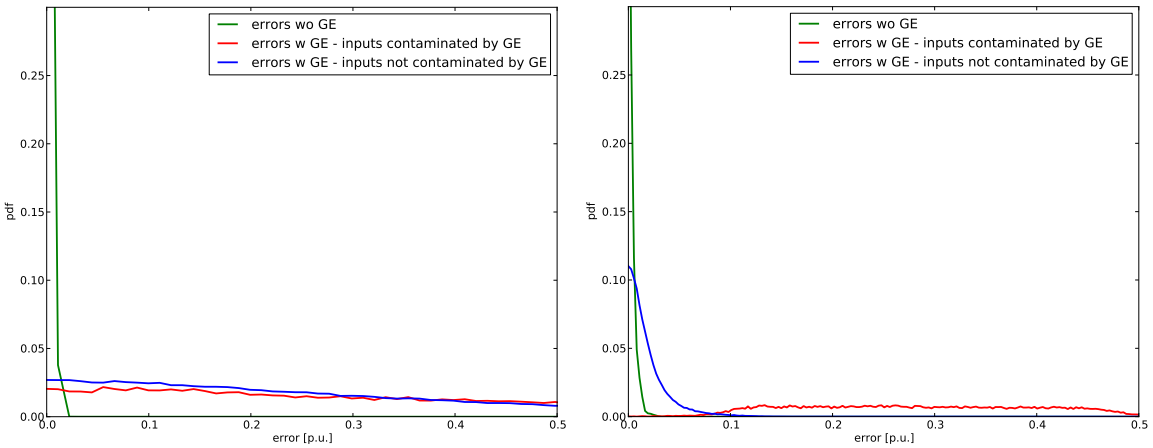


Figure 4.8: Scatter plot of absolute input-output errors per each input for the denoising autoencoder. It is important to note a significant enhancement when comparing the same plot for normal autoencoder shown in Figure 4.6. Errors of non-corrupted inputs are low, while the erroneous inputs also correspond to outliers among the errors.

Figure 4.9 shows PDF of errors for the denoising and normal autoencoder. When comparing Figures 4.9(a) and 4.9(b), the following properties that reveal superiority of the denoising concept can be observed:

- PDF of errors contaminated by GE (red line) is clearly distinguishable from the PDF of errors corresponding to inputs free of GE (blue line),
- In case of correct inputs (without GE), the input-output errors of DAANN are on much lower level, the same as in case of normal AANN.



(a) Normal Autoencoder

(b) Denoising Autoencoder

Figure 4.9: PDF of absolute errors for non-corrupted and corrupted test datasets: (a) Normal Autoencoder; (b) Denoising Autoencoder. The denoising autoencoder obviously outperforms the normal one, since it clearly distinguishes PDF of errors for erroneous and true inputs.

It can be concluded that if the input vector is free of GE, DAANN performs the same as AANN. On the other hand, when some of the inputs contain GE, DAANN significantly outperforms the normal AANN, since it manages to project the outlier back onto the proper place on the manifold. Therefore, the denoising training concept allows accurate manifold learning, and contributes with significant robust features in the meantime, that seem very promising for bad measurement screening.

4.2.4 Gross Errors Detection, Identification and Correction with Denoising Autoencoder, MCC and EPSO

In the previous section a substantial enhancement in robust features provided by the denoising training concept has been shown. However, the identification efficiency still fails to reach a performance level considered to be acceptable in state estimation. In order to extract maximum information stored in autoencoder parameters, the autoencoder is proposed to be coupled with an optimization method to be able to fine-tune the inputs and project them as close as possible to the true point on the manifold. In other words, an optimization method will search for the optimum of the objective function of errors between the output and the available measurement vector. The concept of such constrained search optimization is visualized in Figure 4.10. There are no fixed inputs and all inputs are considered to be missing instead. The optimization method is expected to estimate inputs so that a function of errors between autoencoder's output $\hat{z}^{AANN}$ and available measurements is minimal. The optimization method adopted for this purpose is again the meta-heuristic method EPSO. The only question left is a definition of the objective function. If the square function is used, as in eq. (4.1), inputs of very poor accuracy are mostly obtained, with

insignificant enhancement in comparison to the denoising autoencoder itself. The reason for this comes from the lack of robust features of the least squares (MMSE) criterion, since the euclidean measure of similarity becomes significantly biased in the presence of GE.

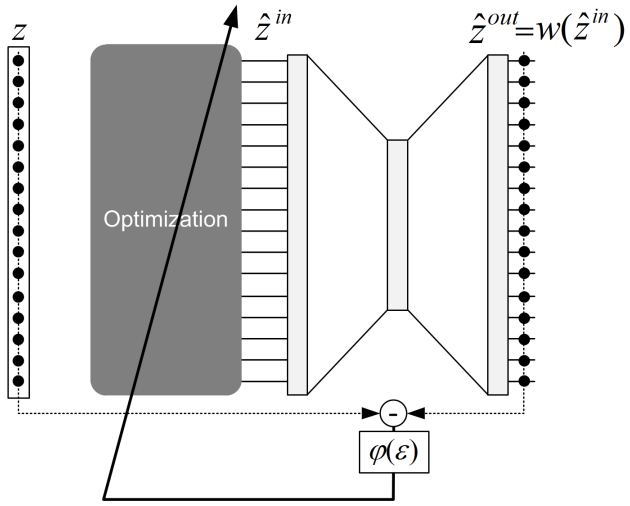


Figure 4.10: Coupling of autoencoder with an optimization method in order to avail the identification and quantification of gross errors. All the inputs are considered to be missing and the optimization method intends to estimate the inputs in order to minimize/maximize the objective function of errors between autoencoder's outputs and the original measurement vector.

Therefore, once again, one proposes to avail the robust features of measures defined within the Information Theoretic Learning (ITL) [113]. The principal idea is to maximize the information that output of the autoencoder provides about the measurements acquired by the SCADA. To implement this, one will minimize the CIM distance (section 3.1.1) between the available measurements and output of DAANN. This criterion is equivalent to maximization of similarity between these two vectors measured with Correntropy:

$$\varphi_{MCC}(\varepsilon) = \sum_{i=1}^{N} e^{-\frac{(z_i - \hat{z}_i^{out})^2}{2\sigma_i}} \tag{4.2}$$

where the free parameter σ represents the size of the Parzen window, which is set to a fix value ($0.02p.u.$), according to the largest error which is still not considered as gross error. Vector $\hat{z}^{out}$ is an output of the denoising autoencoder and ε is an error vector between the input measurements and the autoencoder's outputs $\varepsilon = z - \hat{z}^{out}$. Unlike the least squares objective function, the MCC with EPSO, allows full employment of the robust features of a denoising autoencoder.

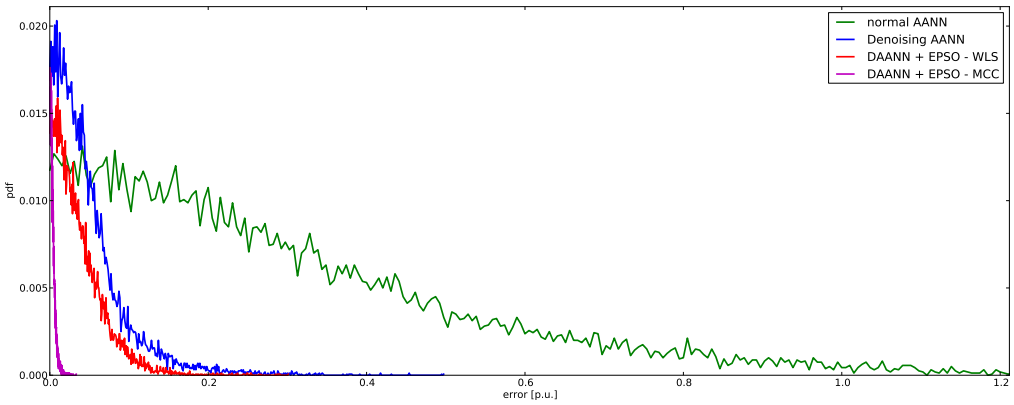


Figure 4.11: PDF of absolute errors from true measurements if input measurements contain 2 GE for all 1000 scenarios. Significant enhancement is achieved by denoising way of training. The best accuracy is by far obtained if EPSO optimization is used to maximize the Correntropy between the output and the original measurement vector.

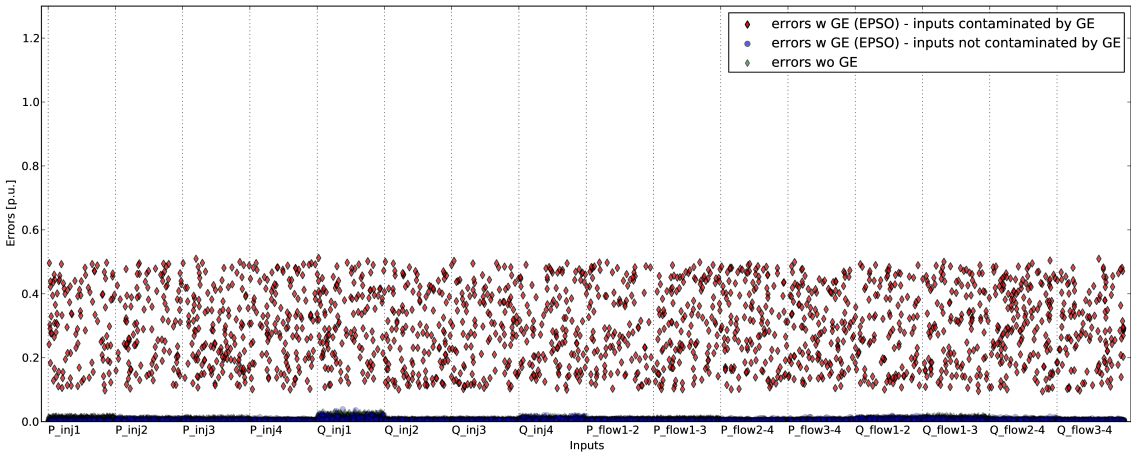


Figure 4.12: Scatter plot of input-output absolute errors per each input for the denoising autoencoder whose inputs were optimized in order to maximize the Correntropy criterion. The desired shape of the errors scatter plot is finally obtained, with significant improvements when compared to the same plots for a normal autoencoder (Figure 4.6) or denoising autoencoder (Figure 4.8) .

The results are presented in Figure 4.11 through the PDF of absolute errors. A distribution of errors for output vectors (obtained with inputs corrupted by 2 GE) from the true measurement vector is calculated for all methods in 1000 scenarios. The worse performance shows the normal autoencoder, which is evidently very sensitive to GE, and error smears for all outputs. A significant improvement in robustness is provided by the concept of denoising autoencoder training. If an optimization method is coupled to fine tune the estimated measurements due to quadratic criterion, there is no significant enhancement of robust features. Finally, if an optimization method maximizes information, the output of the denoising autoencoder provides about the

available inputs, the estimated values get a substantially improved accuracy in comparison to true measurements. This is obtained by implementation of MCC criterion, which together with EPSO and denoising autoencoder provides an impressive robustness to multiple GE. This method of measurements screening is denoted hereafter as DAANN-EPSO.

Scatter-plot of absolute errors for DAANN-EPSO from Figure 4.12, finally gets the desired shape. Now, errors of inputs which are not contaminated by gross errors are reduced to close to zero, the same as errors from inputs without any GE. On the other hand, all errors from the contaminated inputs (red points) are in the expected range of GE $[0.1, 0.5]$, which means that the proposed method is able to identify and, to a certain extent, also quantify the gross errors.

This is a substantial advancement in comparison to same plots for the denoising autoencoder (Fig. 4.8), or even more when compared to the normal autoencoder (Fig. 4.6). Not only does the proposed method manage to correctly identify bad data with 100% efficiency, but it also corrects the erroneous measurements with remarkable accuracy. In regard to the significance of these results, the basic idea of this chapter is to use autoencoders as measurement screening processors prior to the estimation of the system state.

The following section addresses all the issues and possibilities of adopting autoencoders in SE procedure as suppliers of correct information about the system in all possible scenarios.

4.3 State Estimation with Denoising Autoencoders and Maximum Correntropy Criterion (DAANN-MCC)

The previous section demonstrates strong capabilities of auto-associative neural networks for screening of any type of electric variables in power systems. Unlike the classical methods, which estimate the system state according to single measurement snapshot, the proposed auto-associative measurement processing extracts the maximum information about the system in accordance with the entire historical dataset. This way, the autoencoder is able to perform robustly and accurately even in cases where classical methods are theoretically unable to detect any of the bad data.

Since autoencoders just filter the measurements and do not provide final estimative of the system state, this chapter proposes a methodology which allows the use of autoencoders as pre-processors. More specifically, a hybrid method is established which couples auto-associative data processor with the MCC estimator (proposed in Chapter 3). Basically, the autoencoder filters the acquired measurement vector and forwards the results to the MCC estimator, together with an appropriate Parzen window scaling. More specifically, the autoencoders suspect bad data and according to this, the Gaussian kernels bandwidth of the Correntropy function will be set. In summary, this section proposes the final state estimation methodology which incorporates all the following specific concepts:

- ▷ Maximum Correntropy Criterion (MCC),
- ▷ Parzen window annealing,

▷ Capturing data manifold with auto-associative neural networks,

▷ Denoising Auto-associative Neural Networks (DAANN),

▷ Data screening with constrained search of autoencoders inputs (DAANN-EPSO).

Hence, the hybrid methodology intends to make use of all valuable features of the adopted concepts, with the main intention to achieve maximum of robustness, but without threating computational feasibility.

The basic idea of coupling autoencoders with MCC estimator is to use the denoising autoencoders as suppliers of the corrected information about the system for the MCC state estimator. During the training process, autoencoders capture some extra knowledge about the power system characteristics, which are generally not available to a static state estimator. This information, captured during the training, is being extracted through the screening process and then provided to the MCC. With this extra information, the MCC robustness is expected to be enhanced by eliminating the failures which are probable to occur in multiple GE scenarios. Moreover, since observability, in the classical sense, is not a prerequisite for the autoencoders, one expects from the proposed method to allow identification and quantification of GE even in critical measurements. On the other hand, classical (static) state estimators are not even theoretically able to identify bad critical measurements since the corresponding residual is always equal to zero. Hereafter, performances of the hybrid method are preliminary tested on multiple case studies. Also, the final hybrid SE algorithm is established.

4.3.1 Preliminary Test Case Study

The 4-bus system is used as a preliminary test sample for the hybrid state estimator. The methodology is basically conceptualized as the following: the MCC estimator (proposed in Chapter 3), instead of the original measurements acquired from the SCADA, uses outputs of the (denoising) autoencoders. The following case studies are considered for a complete assessment of the method performances:

1. multiple interacting and conforming GE,
2. 1000 scenarios of 2 GE with high measurements redundancy,
3. 1000 scenarios of 2 GE with extremely low measurements redundancy,
4. bad critical measurement.

All the measurements scenarios are generated by adding zero-mean Gaussian errors (with variance $\sigma = 0.005 p.u.$) to true measurements computed using a power flow solution. The following measurement vectors (composed of active and reactive power injections and flows and voltage magnitudes) are used as inputs for the MCC state estimator:

– original measurements free of GE,

– original SCADA measurements contaminated with GE,

– output from DAANN,

– output from DAANN-EPSO.

Test case 1: multiple interacting and conforming GE

The first test case represents an extreme gross errors scenario with 4 GE in the 4 injection active power measurements with the following values $\{-1, 1, -1, -1\}$, respectively. A bar chart of residuals, associated to all used methods, is presented in Figure 4.13. As expected, a strong smearing and masking effect is obvious in the WLS result. The LNRT method fails to identify all the gross errors due to a masking effect that conforming GE cause among the residuals. On the other hand, the denoising autoencoder (DAANN) itself (without EPSO) obtains much better result and identifies correctly all the 4 GE. However, a certain lack in precision is obvious. Finally, when EPSO is used to maximize the Correntropy between autoencoder outputs and the original measurements (DAANN-EPSO), all 4 GE are quantified with almost perfect precision. This corrected measurements set is then forwarded to the MCC state estimator which provides the state vector which corresponds to the true one (DAANN-MCC).

This is quite an extreme scenario with 4 incident and conforming GE in the leverage points, which are able to deteriorate any classical state estimation. On the other hand, the proposed hybrid state estimator based on autoencoders and Correntropy criterion manages to successfully identify all GE and even correct them with remarkable accuracy.

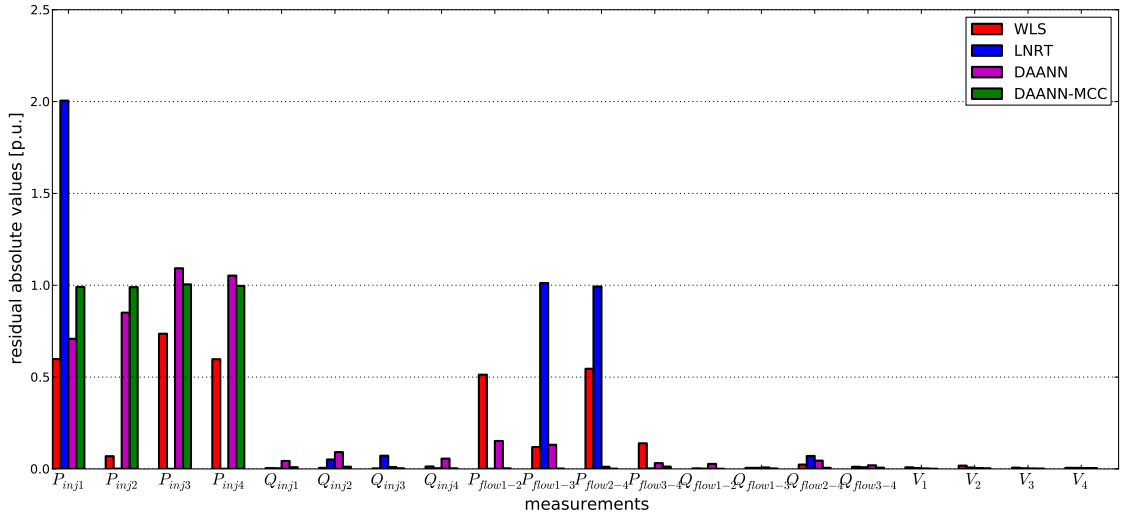


Figure 4.13: A bar chart plot of absolute residuals of 4 methods: WLS, LNRT, DAANN, DAANN-MCC in the presence of 4 GE. Only DAANN-MCC manages to identify and correct them with a satisfactory precision.

Test case 2: 1000 scenarios with 2 GE - high redundancy

In order to derive a statistically more sound conclusion, 1000 measurements scenarios corrupted with 2 GE are considered. A true measurement vector, (without GE and without the noise), is taken

as a reference for calculation of estimation errors in order to evaluate the estimation results. For this case study, a high measurements redundancy is assumed (2.5): 8 power injections (4 active and 4 reactive power), 8 power flows (4 active and 4 reactive power) and 4 voltage magnitudes are considered.

The PDFs of absolute errors are shown in Figure 4.14. It can be clearly seen how GE strongly deteriorate error PDF for the WLS and DAANN (without EPSO). On the other hand, DAANN-MCC results with PDF practically equivalent to the error PDF without GE. This is a very important result since it demonstrates high immunity of the proposed method to any kind of multiple GE. The measurement screening process with DAANN and EPSO is extremely precise and accurate for adoption in any state estimator as a provider of high quality data, purified from any *non-normal* errors.

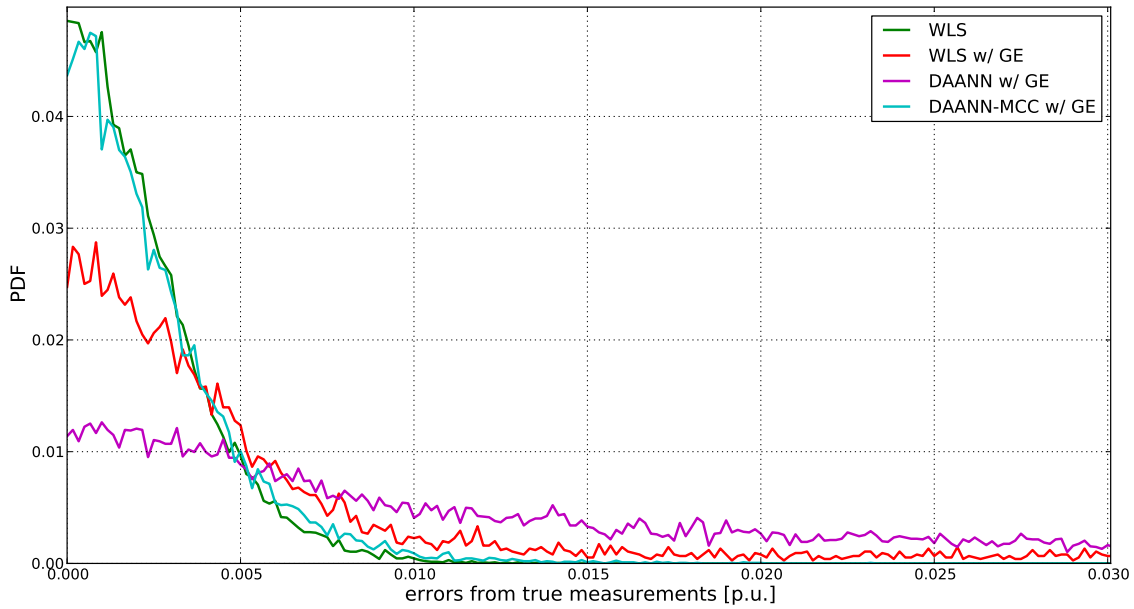


Figure 4.14: PDFs of absolute estimation errors (from true measurements) for WLS, DAANN or DAANN-MCC methods. If scenarios with and without the GE are compared, a significant deterioration of results is obvious for both WLS and DAANN. Only errors of the hybrid DAANN-MCC method seem to remain on the level of introduced Gaussian noise.

In order to examine the obtained results in more detail, the efficiency in GE identification is evaluated. The conventional Largest Normalized Residual Test (LNRT) method is also included for comparison.

Identification is considered to be correct if a measurement set, suspected as erroneous, corresponds exactly to introduced gross errors. In case of any misidentification, a failure is recorded. The results are represented in Table 4.1. It is shown once more that LNRT and MCC estimator are able to misidentify GE, especially if they are interacting and conforming. A significant novelty brought by these results is: if the MCC estimator is equipped with a data pre-processor, based on

artificial intelligence tools dealing with Correntropy, robustness to GE is substantially enhanced. In summary, the hybrid DAANN-MCC method has successfully identified GE in all scenarios.

Table 4.1: Efficiency of 2 GE identification in 1000 scenarios. Unlike the LNRT and MCC the DAANN-MCC method is able to correctly identify all GE, even in the incident and conforming cases.

	Wrong	**Correct**	**Efficiency**
LNRT	61	939	93.9%
MCC	79	931	93.1%
DAANN-MCC	0	1000	100.0%

Test case 3: 1000 scenarios with 2 GE - poor redundancy

In the former test case, the denoising autoencoder has shown impressive performance, significantly outperforming the conventional methods. This test case further examines performances of the hybrid method in conditions of very poor redundancy level. Even more beneficial performances of this method are particularly expected in cases when only a few measurements are available.

In comparison to the former sample, measurements available in the 4 bus system are reduced from 20 to just 10: P_1, Q_1, P_4, Q_4,P_{1-2}, Q_{1-2}, P_{3-4}, Q_{3-4}, V_1 and V_4. Accordingly, the redundancy level is decreased from 2.5 to 1.25. In general, with low level of local redundancy more measurements behave like leverage points, so a significant efficiency drop is expected for LNRT and MCC.

As expected, both the LNRT and the MCC result in a high GE identification efficiency (Table 4.2). On the other hand, the DAANN-MCC remains impeccable despite a substantial drop in redundancy.

Table 4.2: Efficiency of 2 GE identification in 1000 scenarios considering poor measurements redundancy. The LNRT and MCC manages to identify correctly GE in circa every fifth scenario, while DAANN-MCC identifies correctly all the bad data.

	Wrong	**Correct**	**Efficiency**
LNRT	787	213	21.3%
MCC	792	208	20.8%
DAANN-MCC	0	1000	100.0%

The static-state estimators, by observing only a single snapshot, receive very poor information since the measurements/unknowns ratio is very weak in this case. Accordingly, a significant robustness to errors cannot be expected. On the other side, during the training process the autoencoder stores in its weight all the extractable information about the data, i.e. Kirchoff laws, load distribution curve, network topology, node character (P-V, P-Q) or any other mutual information between inputs. Afterwards, during the reconstruction phase (EPSO maximizes

Correntropy function between autoencoders outputs and original measurements) the information is backed-up in relation to acquired measurement vector. The MCC is then run with the novel vector provided by the autoencoder and with the novel Parzen windows scaled in accordance with the suspected measurements. This extra information conveyed by the data screening process enables a significant enhancement of the MCC robustness.

Test case 4: bad critical measurement

The most demanding scenario in GE identification is when they appear in critical measurement. For classical state estimators which optimize some function of residuals, even detection of such bad data is impossible due to the fact that critical measurements always have residuals equal to zero. Abilities of the proposed auto-associative data screening are tested here in this extreme GE scenario.

Meters within the 4-bus system are again reduced in order to create some critical measurements. The considered measurement set is: P_1, P_{1-2}, P_{1-3}, P_{3-4}, Q_1, Q_{1-2}, Q_{1-3}, Q_{3-4} and V_1. A single GE of $2p.u.$ is introduced in the critical measurement. The detailed results of the WLS and DAANN-MCC estimation are shown in Table 4.3. The WLS Residual of the bad measurement is equal to zero and all other residuals are low, thus making this single GE undetectable using the χ^2-test or normalized residual test. The DAANN-MCC method once again proves remarkable robust capabilities. It identifies and very precisely corrects the bad critical measurement. Other measurements are also very accurately estimated, so the GE is very efficiently isolated and eliminated from the input set.

Table 4.3: State estimation results considering bad critical measurement P_{3-4}. Residual of the critical measurement is equal to 0, so GE is not even detectable. On the other hand, the DAANN-MCC manages to identify and impressively accurately correct the bad critical measurement.

measur.	z^{true}	z^{bad}	$\hat{z}_{WLS}$	r_{WLS}	$\hat{z}_{DAANN-MCC}$	$r_{DAANN-MCC}$
P_{inj1}	0.63003	0.63003	0.62465	0.00538	0.62538	0.00465
$P_{flow1-2}$	0.04187	0.04187	0.04724	-0.00538	0.04633	-0.00446
$P_{flow1-3}$	0.57203	0.57203	0.57741	0.00538	0.57906	-0.00702
$P_{flow3-4}$	-1.09460	0.90540	0.90540	**0.00000**	-1.09546	**2.00087**

An additional, more exhaustive extreme test is performed with 1000 scenarios of the 4-bus system, considering only critical measurements (redundancy level = 1). The available measurements are: P_{1-2}, P_{1-3}, P_{3-4}, Q_{1-2}, Q_{1-3}, Q_{3-4} and V_1. In each scenario, a single GE is introduced in the critical measurement $P_{flow3-4}$. The LNRT and MCC state estimators are not able to perform any robust function in this case. However, the abilities of DAANN-MCC are once again impressive: a successful identification of gross error in 98.7% of cases.

These results are very significant and promising, indicating an enormous potential of the proposed method, particularly in environments like distribution grids, where lack of monitoring is

inevitable.

4.3.2 Method Decentralization - Mosaic of Local Autoencoders

The proposed hybrid method which couples denoising autoencoders with EPSO optimization and MCC criterion (DAANN-MCC) shows impressive robust properties in test cases from the previous section. The only concerning issue, left undressed relates to applicability of the method for more realistic systems of larger size. For the 4-bus system, a denoising autoencoder has considered all the available measurements. If this model is scaled to a large system, the autoencoder may become enormous, implying the following two main concerns:

- ▷ exponential growth of weights number with the autoencoder size,
- ▷ large number of unknowns that need to be optimized by a constrained search.

The first issue is not a big concern since the training process is performed off-line. However, one believes that precision and flexibility of an autoencoder may become seriously harmed with its size. A factor that can be more limiting is the constrained search with EPSO in a large dimensional space. More specifically, since the data screening is envisaged to be performed in real time, the computational requirements may become an insuperable burden.

A solution for the method scalability proposed in this section is in decentralization of approaching the problem. Instead of a single global autoencoder, the concept of ***mosaic of local autoencoders*** is established. Justification of the concept comes from the fact that any operational change in an electric power system is evident mostly in some relatively narrow local area and further vanishes by increasing electrical distance. Therefore, local autoencoders of reduced size, are spread over the system to be in charge of associated local zones in the system. In this concept, the superposition among autoencoders is preferable since a certain drop in precision is expected towards the boarders of the local zone.
Benefits of this decentralized architecture are multiple. Firstly, the number of weights that need to be trained in local autoencoders is drastically reduced. Also, the flexibility of the method is enhanced in regard to re-training of the autoencoder in the cases of system structural changes.
When some of the local autoencoders alarms the presence of gross errors, they are automatically localized to a small area. Therefore, it is sufficient to perform a constrained search only for the local variables thus drastically reducing the computational demand. Moreover, a decentralized concept allows for all the calculations, from training to optimization, to be performed independently and in parallel. The decentralization method is expected to be beneficial even due to GE processing, since the probability of an autoencoder to receive a corrupted input is significantly reduced. In addition, the superposition of autoencoders can bring a certain redundancy in GE processing.

The proposed concept of mosaic of local autoencoders is tested on the 24-bus IEEE system [130]. Figure 4.15 depicts the areas defined for the 7 local autoencoders, from which they collect the available measurements of active or reactive power. All the 7 autoencoders were trained on the denoising way, with sizes from 10 to 14 inputs of active or power measurements.

Redundancy in the system was relatively high, considering 24 injection measurements (active and reactive), 34 flows (active and reactive) and 4 voltage magnitudes.

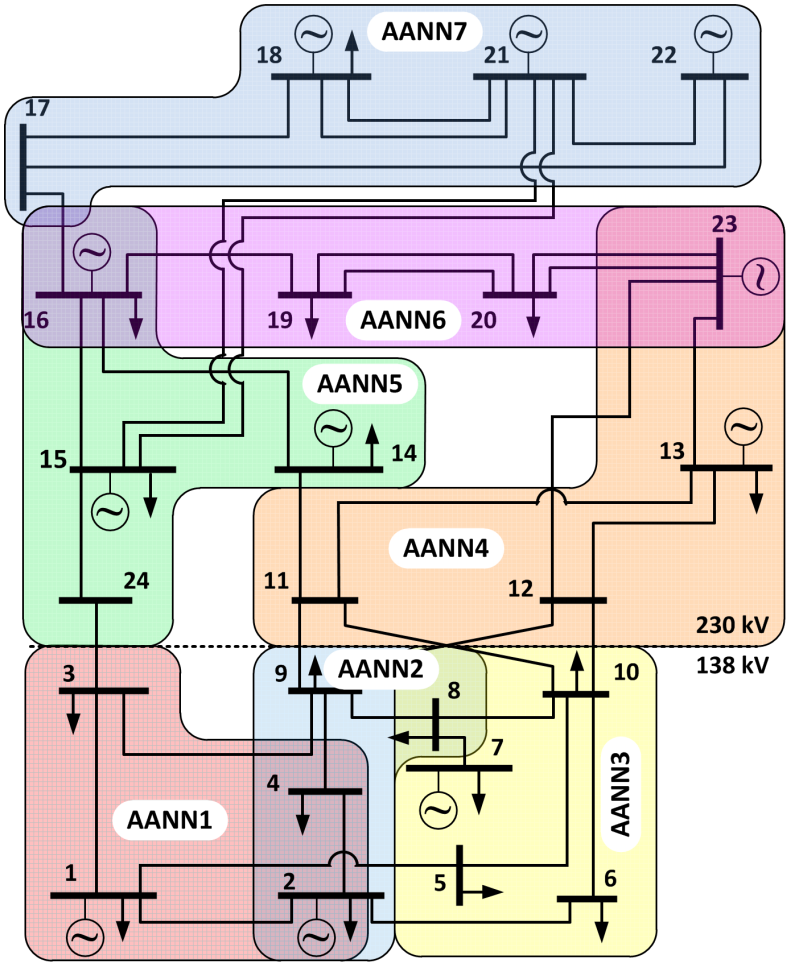


Figure 4.15: Mosaic of 7 local autoencoders in the 24-bus system. Each autoencoder has some local area which it is in charge of, including the zones of superposition.

The first test case assumes 5 gross errors in the following injective power measurements: $P_3, P_{24}, P_{21}, P_{22}, P_{23}$ with the following values $\{1,-1,1,-1,1\}$, respectively.
Since multiple conforming bad data are included, the LNRT and the MCC fail to identify GE due to the masking effect. After submitting the available measurements to the local autoencoders, 5 of them (AANN1, AANN4, AANN5, AANN6 and AANN7) flagged the presence of gross errors in their inputs. Then the EPSO optimization is run to maximize the Correntropy function and the following results are obtained:

- AANN1 identifies GE in P_3
- AANN4 identifies GE in P_{23}
- AANN5 identifies GE in P_{24}

- AANN6 identifies GE in P_{23}
- AANN7 identifies GE in P_{21} and P_{22}

Therefore, all of the 5 introduced gross errors are correctly identified. In addition, due to superposition among the autoencoders, the measurement is identified twice by autoencoders AANN4 and AANN6. The estimation results for a particular scenario are presented as a PDF of errors from true measurements in Figure 4.16. PDFs are estimated by the Parzen window estimation method, using Gaussian kernel with size of $\sigma = 0.005$. One can observe how GE had significantly deteriorated the results of the WLS and LNRT. On the other hand, errors of values estimated by DAANN-MCC are all clustered around zero. Therefore, bad measurements are not only identified but also quantified and corrected. This is a key difference from the LNRT method, which only eliminates identified measurements and thus reduces redundancy and threatens system observability.

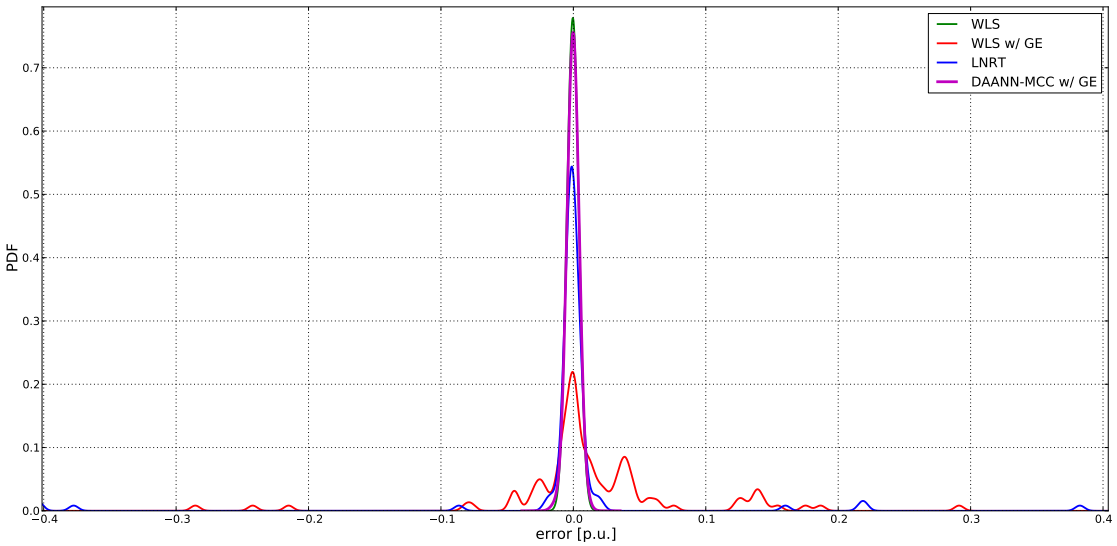


Figure 4.16: PDFs of errors estimated by the Parzen window method using $\sigma = 0.005$. The WLS and LNRT are significantly harmed by GE, while DAANN-MCC successfully identifies and corrects 5 GE.

The computational performance the method has in this particular case, required approximately the same time as the LNRT to process bad data and provide the final solution. Although these methods are barely comparable, due to significant difference in their concepts, it is important to show that the DAANN-MCC method is scalable to larger problems. Outputs of an autoencoder are obtained practically instantaneously, since it only requires evaluation of a certain number of simple algebraic functions. Therefore, the detection process is straightforward, while the identification and quantification of GE requires an optimization process whose computational costs are never expected to be abundant. The main computational cost pertains to the constrained search, in this case to the meta-heuristic method - EPSO. With the concept of local autoencoders, a constrained

search is run only in case of GE detection and only in small areas where they are localized. This way, a number of variables considered by EPSO is substantially reduced. In summary, the proposed decentralized estimation architecture is expected to avail significant robustness from the auto-associative data screening in real time.

4.3.3 DAANN-MCC State Estimation - the Algorithm

In former sections, a novel state estimation methodology is proposed and preliminarily tested on multiple case studies. After the acquisition of valuable conclusions about the method performances, this section proposes the final algorithm that efficiently couples the MCC estimator (proposed in Chapter 3) and the auto-associative measurement screening process. The main intention of the algorithm conception is to keep all of the demonstrated robust features of autoencoders without sacrificing the precious computational resources.

The basic algorithm flow is given hereafter, followed by a principal illustration shown in Figure 4.17. The basic idea of the algorithm is to selectively employ the computationally most demanding auto-associative data screening in cases when it is necessary. Also, not to consider the entire system but only the local zones suspected to be corrupted by GE. The core of the method is the MCC estimator which is computationally efficient and also reliable in most of the cases. In relatively rare scenarios when the MCC estimator failure is suspected, the decentralized data screening process is activated. After the identified bad measurements have been corrected, they are submitted again to the MCC to re-estimate the state.

However, it is not necessary to consider now the entire system, if knowing that the effect of GE in a power system remains limited locally, i.e. inside a relatively small sub-system. In other words, using the fact that GE propagate poorly in state estimation result, computational requirements of the method may be substantially reduced by performing the re-estimation only on to the so-called bad data pockets. This property becomes of particular importance as the size of the power system increases. Hence, according to bad data suspected from the global MCC, a bad data pocket can be defined as a sub-network for which it is guaranteed that contains a GE.

Definition of data pockets is performed according to measurements/state adjacency algorithm proposed in [92]. Specifically, using the Jacobian matrix structure, state variables adjacent to the suspect measurement set are selected. Afterwards, following again the Jacobian matrix structure, suspected measurements set is extended by those measurements which are functions of the selected state variables. Finally, the bad data sub-network is formed by the final set of suspected measurements and states. In case multiple pockets are overlapping, they will be automatically merged in a single bad data pocket.

In these critical scenarios, when MCC failure is suspected and autoencoders are activated for measurement screening, the estimation process will lead to two possible solutions: one from the initial global MCC estimation, and the other one from the local MCC which considers only bad data pockets. Among these two, the final solution is selected according to the larger value of the Correntropy function. This is justifiable by the fact that global maximum of the Correntropy function always corresponds to the true solution (unless in case of bad critical k-tuple).

Algorithm 2 DAANN-MCC state estimator algorithm (fig. 4.17)

GE detection by processing all the available measurements with the mosaic of local autoencoders
Topology processing and observability analysis
Global MCC state estimation - results with $\hat{x}_{MCC}$, $\hat{z}_{MCC} = h(\hat{x}_{MCC})$
if MCC or any autoencoder detected GE **then**
    Use the measurement/state adjacency algorithm to build a bad data pocket around each suspected measurements
    if MCC identified a single GE within each of the bad data pockets **then**
        $\hat{x}_{final} = \hat{x}_{MCC}$
        go to END:
    Perform bad data screening on the bad data pockets which results with the corrected measurements vector $\hat{z}_{DAANN-EPSO}$.
    if GE identified by MCC S_{MCC} are identical to GE identified by DAANN S_{DAANN} **then**
        $\hat{x}_{final} = \hat{x}_{MCC}$
    else
        Perform local MCC estimation on the bad data pockets using the relevant $\hat{z}_{DAANN-EPSO}$ and appropriate Parzen window scaling which results with $\hat{x}_{DAANN-MCC}$
        if $Correntropy(\hat{x}_{MCC}) > Correntropy(\hat{x}_{DAANN-MCC})$ **then**
            $\hat{x}_{final} = \hat{x}_{MCC}$
        else
            $\hat{x}_{final} = \hat{x}_{DAANN-MCC}$
END:

There are some important properties of this method which deserve additional attention. Firstly, in the scenarios free of GE, the proposed method will perform equivalently to the WLS method. Detection of GE is performed by a mosaic of autoencoders, which enables detection of eventual bad critical measurements. Since the MCC always succeeds to identify a single GE or multiple non-interacting GE, one has a reason to doubt the MCC solution only when multiple interactive GE occur. Therefore, after the gross errors have been identified by the MCC estimator, the local zones of interactive measurements are defined. If only a single GE is identified within each zone, the GE are non-interactive so MCC is not prone to failure and its solution can be accepted as the final one. On the other hand, if multiple interactive GE are identified by the MCC, the estimation process focuses on a specific local zone and an auto-associative screening is performed. After the suspected bad measurements are corrected, they are submitted to the second MCC estimator which is run locally, only considering measurements and state variables inside the local area where the misidentified GE may be located, i.e. inside the bad data pockets. Moreover, if autoencoders identified the same GE as the global MCC, the local MCC estimation is not necessary and the state vector estimated by the first MCC is accepted as the final solution.

In the following sections, performances of the proposed algorithm are tested and compared with other relevant estimators in multiple representative case studies.

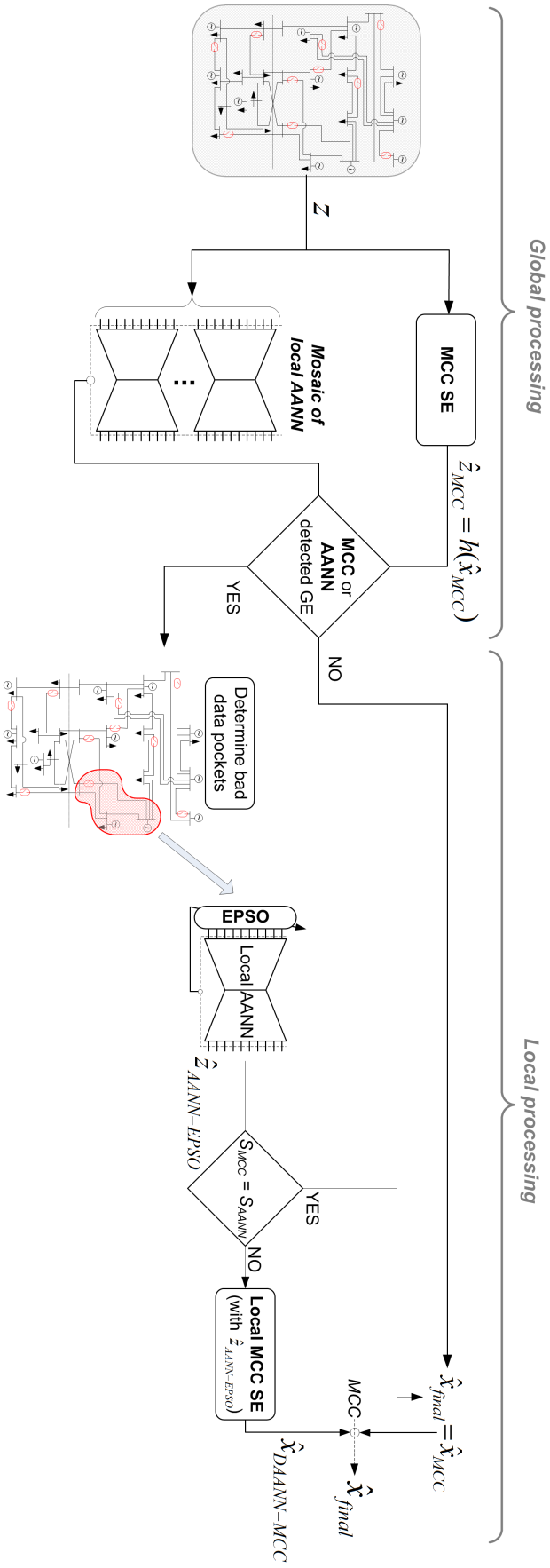


Figure 4.17: Scheme of the proposed DAANN-MCC state estimation algorithm.

4.3.4 Testing of the Method Performances

The algorithm for the proposed DAANN-MCC estimator is tested in this section on the three representative test systems. Once again, the results were confirmed on a small 4-bus system in comparison to all other state estimators analyzed through this thesis and a more exhaustive testing is performed on the 24-bus system. Finally, the 118-bus IEEE system is considered, as a large-enough network to prove the scalability of the method.

4.3.4.1 Case Study - 4-bus System

Performances of the proposed hybrid state estimation method are tested in 10000 scenarios on the 4-bus system (more about the system in Appendix A.1). Each measurement scenario contains 2 randomly generated gross errors only in active power measurements (in order to increase the frequency of interacting and conforming GE occurrences). The computational requirements are not analyzed in this case study since they are not concerning.

The following key remarks are derived from results on efficiency in GE identification, reported in Table 4.4:

- The pure static estimator, LNRT and MCC are prone to failure when multiple GE are incident and conforming. In other cases, which are the vast majority, they manage to successfully identify GE and eliminate their distorting effect from the final solution.

- The proposed DAANN-MCC estimator successfully identifies GE in all the 10000 scenarios, i.e. unlike the LNRT and the MCC, it is able to identify even multiple conforming GE.

- The superiority of the DAANN-MCC estimator is considerable in one more ability of key importance: unlike the static estimators which eliminate the bad data, this estimator is able to quantify the GE and correct the measurement vector without deteriorating the redundancy level.

Table 4.4: Efficiency in identification of 2 GE in 10000 measurement scenarios of the 4-bus system. The LNRT and MCC estimators fail to identify incident and conforming GE, while DAANN-MCC outperforms them with impeccable performance.

	Wrong	**Correct**	**Efficiency**
LNRT	641	9359	93.59%
MCC	1169	8831	88.31%
DAANN-MCC	0	10000	100.00%

4.3.4.2 Case Study - IEEE 24-bus System

The novel hybrid state estimation method is tested here on a significantly larger system - the 24-bus IEEE RTS system (Appendix A.2). The 10000 measurement scenarios are generated as the test dataset, including 2 GE, generated only in incident active power measurements. The special focus in the performance analysis is put on the computational requirements, since they are a prerequisite for this network size.

Table 4.5: Efficiency of identifying 2 interactive GE in active power measurements considering 10000 measurement scenarios of the 24-bus system. The proposed DAANN-MCC estimator is impeccable in GE identification, unlike the LNRT and MCC estimators which are misidentifying incident and conforming GE scenarios.

	Wrong	**Correct**	**Efficiency**
LNRT	783	9217	**92.17%**
MCC	1524	8476	**84.76%**
DAANN-MCC	0	10000	**100.00%**

The GE identification efficiency results from Table 4.5 are confirming the results obtained in the previous case study (the 4-bus system). It is more important however to observe the computational performance results presented in Tables 4.6 and 4.7. Hence, it can be concluded that this remarkable robustness is not achieved due to some unreasonable computational requirements. Most of the required computing time is spent on the auto-associative data screening with DAANN and EPSO. According to this, it is important to note that the represented results correspond to the most demanding scenarios, when GE are incident. If non-incident cases were included in the test, the computational costs would be much lower, since the EPSO search is not activated in these cases.

Table 4.6: Computational performance results for state estimation of the 24-bus system with 2 GE. The computational time of the hybrid estimator (DAANN-MCC) is larger then MCC at the cost of computational force spent on the auto-associative measurements screening (DAANN-EPSO). This is the worse possible computational performance of the DAANN-MCC estimator since only incident GE are considered.

	mean(t_ω) [s]	std(t_ω) [s]	min(t_ω) [s]	max(t_ω) [s]
WLS	0.1261	0.0233	0.0910	0.2000
LNRT	0.5088	0.1640	0.2610	1.8270
MCC	0.4837	0.1099	0.3090	0.9300
DAANN-EPSO	0.7004	0.2989	0.3400	1.4550
DAANN-MCC	1.3048	0.6104	0.7983	2.2421

The results of the other end-case test sample, free of GE (Table 4.7), are also very indicative in understanding the computational performances of the proposed method. Hence, in a case free of GE, the estimation process is effectively reduced to a classical WLS estimation, which is in that case also the most preferable solution. Autoencoders are used in these cases only in the detection process, which is straightforward, while data screening by EPSO is never run. This is obvious from negligible computational requirements of DAANN-EPSO whilst the DAANN-MCC method is on the same level of order as the WLS.

Table 4.7: Computational performance results for state estimation of the 24-bus system, if measurements are free of GE. It is significant to observe that DAANN-EPSO processing employs negligible computational force since it only performs detection, while the screening process is never run. The computational effort of the DAANN-MCC estimator is on the same level of order as the WLS.

	mean(t_ω) [s]	std(t_ω) [s]	min(t_ω) [s]	max(t_ω) [s]
WLS	0.1239	0.0199	0.0960	0.1750
LNRT	0.1652	0.0221	0.1250	0.2230
MCC	0.1638	0.0258	0.1200	0.2330
DAANN-EPSO	0.0400	0.0040	0.0350	0.0550
DAANN-MCC	0.2071	0.0348	0.1590	0.3950

In summary, the proposed hybrid estimation method will be as computationally demanding as the MCC estimator in most of the scenarios. Only in less probable cases, when failure of the MCC is suspected, additional computational costs are expected. However, thanks to the decentralized method architecture, an increase in time requirements for solving these demanding cases will be relatively modest since it does not depend on size of the system, but rather on the number of identified GE and their position in the system. The following case study should confirm the scalability of the method for a larger system.

4.3.4.3 Case Study - IEEE 118 Bus System - Proof of Scalability

In order to prove scalability of the proposed hybrid state estimator, based on Correntropy and autoencoders, in this section it is tested on a large-enough network - the IEEE 118-bus system. The testing database includes 100 measurements scenarios with 2 GE generated among the incident active power measurements.

The results of GE identification efficiency, reported in Table 4.8 repeat the conclusions derived from smaller systems. All the failure cases of LNRT and MCC, which mostly occur when interacting GE are conforming, are all successfully solved by the DAANN-MCC estimator, demonstrating impeccable performance.

More significant results in this case study relevant to computational performances of the methods are presented in Table 4.9. Firstly, it is important to emphasize that the reported results

are considering only incident GE. In other words, these result represent the worst possible computational performance of the DAANN-MCC method. Since in these cases, the MCC estimator is at a larger risk of failure, the method employs measurements screening procedure based on autoencoders and meta-heuristic EPSO optimization. As can be observed from results, most of additional computational costs are spent on data screening (DAANN-EPSO). It is of key importance that these costs are not increasing by the network size, but keep on the same level of order thanks to the localized scheme. Some extra time may be consumed by the local MCC but negligibly comparing to the global MCC. Moreover, local data screening is performed only in cases when bad data suspected by the global MCC and the autoencoders are not the same. In all other cases (non-interactive GE, single GE, or without GE), the screening process is not performed so the method ends with a common MCC estimation.

Table 4.8: Efficiency of identification of 2 interactive GE in active power measurements considering 100 measurement scenarios of the 118-bus system.

	Wrong	**Correct**	**Efficiency**
LNRT	9	91	91.00%
MCC	16	84	84.00%
DAANN-MCC	0	100	100.00%

Table 4.9: Computational performances for SE of the 118-bus system. If the results for the 24-bus system 4.6 are observed, computational requirements of the DAANN-EPSO screening remains approximately on the same level of order.

	mean(t_ω) [s]	std(t_ω) [s]	min(t_ω) [s]	max(t_ω) [s]
WLS	0.7219	0.0875	0.5530	0.9080
LNRT	2.8221	0.4422	2.2860	4.4500
MCC	2.2642	0.8749	1.2570	4.1310
DAANN-EPSO	0.9584	0.3897	0.4280	2.1223
DAANN-MCC	3.9129	1.3195	2.3521	5.571

From the results obtained in this case study, it can finally be concluded that all concerns regarding the scalability of the proposed hybrid DAANN-MCC state estimator and its feasibility in the real time, are eliminated. This is mostly achieved by decentralized architecture of autoencoder mosaic and selective activation of the measurement screening process. In other words, the computational time of the proposed hybrid method will be larger than the computational time of the common MCC estimator only in relatively rare cases, when the MCC failure is suspected and a modest amount of extra-time is sacrificed for the processing that contributes with extreme robustness in any possible GE scenario.

4.4 Conclusions - Auto-associative Information Theoretic SE

The MCC estimator, as proposed in Chapter 3, obviously did not guarantee reaching the maximum of the Correntropy function. The Correntropy function was proved to possess high robust properties, but the solver method that was used (Newton's method + Parzen window annealing) did not ensure finding the global optimum in all the scenarios.

Under the information theoretic learning perspective, this chapter proposes an unconventional method that enables a detachment from the classical approach of mathematic regression and offers an unbiased reference for the MCC estimator. The technique is built around a main tool, constituted by auto-associative neural networks (or autoencoders), able to learn the supporting manifold of the solutions of the power flow equations coupled with other information. After an offline training process, the trained autoencoder is used in real time to screen the reported measurements so that the erroneous inputs are corrected and the eventually missing input signals are reconstructed.

Data inputs in the autoencoder are projected to a different space (usually with reduced dimension, at the hidden layer level) and re-projected back to the original one. If the input is coherent with the manifold learned during the training process, it is re-projected back to the same surface or its close vicinity (to account for small noise and training error). If not, a large input-output difference should become evident. This allows the detection of outliers and provides an attempt of correction of erroneous values (including missing values).

However, this process, taken in a raw manner, also cannot achieve the sufficient precision in the projection process. Therefore, this chapter showed that it is possible to adopt a denoising procedure for autoencoder training that substantially improves their capabilities for GE identification. Furthermore, in order to enable a precise correction of the corrupted data, autoencoders were coupled with an optimization process thus maximizing Correntropy, this time between the output vector and the measurement set acquired by the SCADA. This causes the information that autoencoders provide to be maximized, which leads to a very precise data screening process. Moreover, the autoencoders learn manifold of the data derived from the entire historical dataset and not only by a single snapshot. This enables the identification of GE even in most demanding scenarios, when the conventional methods are not even theoretically capable to propose an acceptable solution. By quantifying gross errors, bad measurements can be precisely corrected, so that observability is not threatened by measurement elimination. In a way, the classical observability requirement is not a pre-requisite for the new method. The lack of classical observability is compensated by extra knowledge about the power system behavior, learned during the training phase.

The main concern about the proposed pre-processing of measurements is the possible computational burden of the technique, which would threaten its feasibility of application in large scale problems. As a solution to this problem, a new concept is offered: the migration from a single and global architecture to a mosaic of distributed local autoencoders. This decentralized architecture of the method drastically reduced computational requirements, and even increased flexibility and robustness.

The contributions provided by the proposed auto-associative decentralized data screening and correction process are summarized hereafter:

- ◇ The adoption of Auto-Associative Neural Networks for measurements screening before being supplied to a state estimator.
- ◇ The adoption of the Denoising training concept in order to enhance the robustness of autoencoders in the identification of GE.
- ◇ The reconstruction of missing measurements and quantification of GE by coupling autoencoders with EPSO in a constrained search. The objective function representing the screening criterion is Correntropy, evaluated between the autoencoder output and the measurement set acquired by the SCADA.
- ◇ The decentralized architecture for the auto-associative data screening that eliminates computational burden in applicability to large scale systems – producing the so-called mosaic. Also, this concept represents an important step towards the integration with the modern power system concepts.

After the testing of all the expected properties of the proposed decentralized methodology for measurement screening and correction, a coupling with the MCC estimator as a new and coherent SE model has been presented in this chapter. This model is built as a hybrid method that makes use of all the valuable features of the MCC criterion in order to achieve an extremely efficient and robust estimation of the power system state. It is a clever blend of information theoretic learning concepts, auto-associative neural networks and the evolutionary optimization, which is taking advantage of distributed architecture concept. In summary, the final contributions of this chapter are the following:

- ◇ A new concept of State Estimation based on distributed architecture and artificial intelligence tools dealing with the information content of data through the Correntropy function has been developed.
- ◇ The novel hybrid state estimator shows extreme robustness to any GE scenarios, even if they occur in critical measurements and therefore it has a superior performance over the most of the existing robust estimators. Also, it is important to note that precious computational force is being selectively and reasonably engaged depending on the detected bad data scenario.
- ◇ The basic trick in the technique is the use of autoencoders as suppliers of a corrected image of the system to the SE procedure – i.e., as if the correction lenses are used before the actual perception process is activated, therefore providing a clearer picture, except for blurring which is due to noise or gross errors.
- ◇ Extreme robust features that include identification and correction of the following most concerning GE scenarios: multiple interacting and conforming gross errors, bad leverage points and bad critical measurements.

Since both the data processor and the state estimator use maximum Correntropy as a criterion, the method is called the hybrid DAANN-MCC state estimator.

One type of gross errors, however, was left out in this analysis and the model proposal: the topology errors. They are induced by erroneous or missing breaker status information. So far, the topology has been assumed to be known and certain. This issue will be addressed in more detail in the following chapters.

Chapter 5

Auto-associative Topology Estimator

Apart from power and voltage measurements, the information about the topology of an electric power system is required in order to carry out a state estimation procedure, or any supervisory and control action in real time operation. The topology is determined by statuses of all breakers existing in a system (i.e., any type of switching devices that influence the network connection), which are in practice represented by binary variables since a breaker can only be in an open or closed state (OFF or ON). Hence, after the breaker statuses are determined for all substations throughout the system, a bus/branch model may be built and a power flow or a state estimation can be run.

The erroneous information about a breaker status may lead to inappropriate or even potentially dangerous control decisions. All the state estimation models discussed in this thesis so far assume a correct power system topology. Such a conventional approach should however be questioned: breaker statuses are provided by signals arriving to the SCADA and, as any other measurement, they are prone to gross errors or may simply be missing due to following reasons: malfunction of the isolation switches, unreported breaker manipulation, mechanical failures of signaling devices, signal transmission failure, etc. In these cases, heuristic rules are often adopted in order to restore the missing value. Alternatively, a breaker status discovery may be attempted with specific topology estimation procedures, which require simultaneous consideration of the breaker status binary variables and the complex non-linear functions that describe power flows. This issue was early recognized and various methods have been proposed for the consideration of topological binary variables within the state estimation procedure.

It is not necessary to emphasize that system topology directly affects the electrical values in a system. In this thesis, a hypothesis that the information about topology lays hidden or is disseminated in values of the analog electric measurements has been formulated. A process of information extraction aimed at topology learning may therefore be conceived to reveal distributed information. Hence, this thesis intends to perceive topology estimation from an information theory point of view, as opposed to most of the conventional mathematical programming methods.

A tool adopted for capturing the information between analog measurements and binary breaker status variables is the Auto-Associative Neural Network (autoencoder). The idea is to

train the neural networks offline in order to learn the manifold implicit in the link between the power flow equations and the topology states. This implies a training procedure sweeping many possible topologies and not only variations on electric values. In the on-line situation, the learned information can rapidly be used to decide on the system topology, taking in account the available measurements. This has been developed in order to achieve a more robust and flexible topology state estimator, independent from any state estimation procedures and able to meet the new challenges of modern power system grids.

The problems addressed, researched and intended to be solved in this section are the following:

- *Is an autoencoder able to learn information about the system topology hidden in the analog measurements?*
- *Is it possible to reconstruct a missing breaker status by extracting the information stored in the weights of a trained autoencoder?*
- *Is the concept of local autoencoders justifiable, i.e. is it possible to locally reconstruct the unknown topology with a performance comparable to the global autoencoder (trained with data from the whole system)?*
- *Is the autoencoder model a feasible alternative to reconstruction of missing switch status signals?*

Most of the research work proposed in this section is already published in the papers [154] and [158]. One may note some difference in the results reported in the thesis for some seemingly same case studies. This is mostly due to the use of a different database, generated for this thesis for the purposes of conformity with all the tests included in the thesis. However, the qualitative conclusions are unaffected.

5.1 Breaker Status Reconstruction with Autoencoders - the Concept Proposal

Most of the conventional topology state estimation methods are based on mathematical programming and require a solution of an optimization problem. A new point of view, advocated in this thesis, relates to information theory and pattern recognition. The rationale behind this is in the fact that electrical variables pertain to a manifold, which is defined by Kirchhoff Laws, load distribution, parameters and characteristics of the system, etc. The shape of this manifold also changes with the state of binary variables related to the topology (i.e., a different topology implies a different solution set). In other words, one can say that values of the topology variables are *hidden* (or equivalently, their information is diffusely included) in the analog electrical variables. Thus, distinct manifolds are associated to distinct topologies.

As already demonstrated in section 4.2, an autoencoder is able to learn the manifold defined by the given measurements. Due to significant results obtained in regard to reconstruction and

filtration of analog measurements, there is no reason to believe that the same approach could not be applied to binary variables of a system topology or a breaker status. Equally to an autoencoder captured in its weights relations between the provided electric measurements, it is also expected to manage to link each measurement vector to a particular topology. In that sense, a simple addition of topology variables to the autoencoder's inputs is suggested. For example, in addition to power measurements, topology binary variables which assimilate the values -1 and 1 for open and closed breaker status, respectively. This way, during the training process, each measurement snapshot is associated to a certain binary code, which represents a particular topology. Values -1 and 1 are chosen as suitable for the *tanh* activation function used in the hidden layer of an autoencoder.

After the autoencoder is trained, it will be used in real time for detection and correction of eventual inconsistencies in the system topology, or for the reconstruction of potentially incomplete topology information. For instance, if one of the breaker statuses is missing, a binary variable is intended to be reconstructed with the constrained search method, (Figure 4.3(c)) which is also used for reconstruction of analog measurements. The optimization criterion is a function of the complete input-output vector distance (error ε). The optimization method used in this thesis is EPSO (the same is adopted in section 4.2 for analog measurements). As previously named in [154] and [158], this method will be further referred to as the self-tuning breaker estimation. Nevertheless, any suitable optimization procedure may be used instead of EPSO: for instance, in [144] a genetic algorithm is employed.

Therefore, when a breaker status signal is missing, the EPSO is used to reconstruct it by searching for an appropriate input value in order to minimize the function from equation (4.1). The status is then obtained by rounding the value to the closest integer value. In order to test the capabilities of such an auto-associative topology state estimator, a preliminary experiment was performed on the IEEE RTS 24-bus system [130]. The system topology is varied via the addition of 10 breakers in the lines as represented in Figure 5.1.

A database used for the training and testing of autoencoders in all the experiments with the IEEE 24-bus system was generated with the following considerations:

a) Insertion of breakers in 10 locations of the network and random, independent status definition for each of the ($2^{10} = 1024$ topologies);

b) Design of a cumulative load curve using data from [130], based on which load levels are sampled with a large variety of scenarios from valley to peak (ratio maximum/minimum power = $100\%/34\%$);

c) Simulation of load variations by adding a Gaussian perturbation with standard deviation $\sigma = 5\%$

d) Generation of a large set of scenarios and power flow calculation results by using the Optimal Power Flow (OPF);

e) Simulation of actual measurements by adding a Gaussian perturbation to power flow solutions, with $\sigma = 0.5\%$ or $0.005 p.u$, ($1 p.u.$ corresponds to 100MVA).

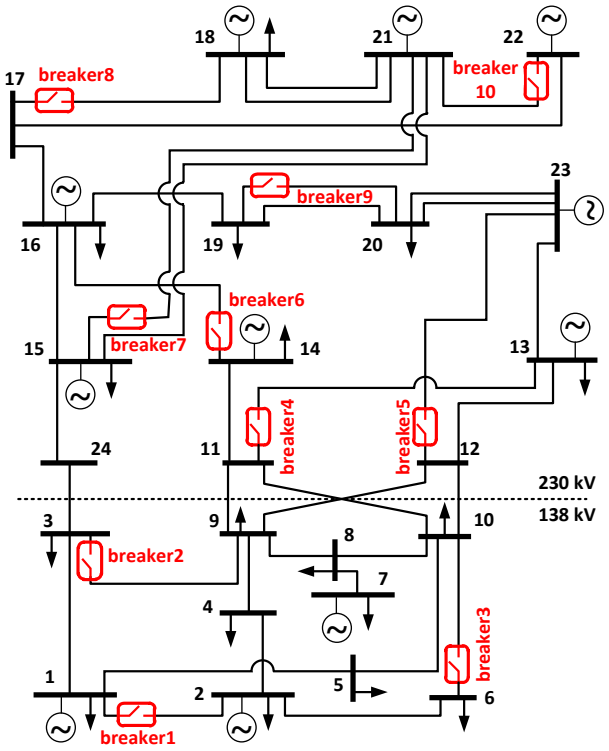


Figure 5.1: IEEE RTS 24-bus system with indication of the branches where 10 switches were introduced.

5.1.1 Global Self-tuning Autoencoder - the Proof of Concept

In order to be able to reconstruct the entire system topology, all the available measurements were used as inputs to the single autoencoder. Thus, it is called the Global autoencoder. Autoencoder's architecture is set to a single hidden layer with an activation function *tanh*, while the output layer is linear. The size of the hidden layer is set to approximately 60% of the number of inputs/outputs. The training of the autoencoder is performed using 10000 measurements vector, generated with random breaker status and load scenarios. The measurement vector is composed of active/reactive power injections in all the nods (2×24), active/reactive power flows measurements in all the lines (2×34) and 10 breaker statuses, what makes in total $2 \times 24 + 2 \times 34 + 10 = 126$ inputs. Obviously, this case study assumed to have a large data redundancy.

After the global autoencoder is trained, the test experiment is conducted by sampling the 10000 scenarios with a diverse active and reactive power measurements and switch status data. For all the 10000 snapshots, 10 sets of trials have been defined, with a fixed number of breaker status signals (1 to 10) proclaimed as missing in each set. Therefore, a total of 550000 of breaker statuses have been reconstructed. It is important to note that system topology is defined by signals from all the 10 breakers. Accordingly, if one of the breaker statuses is wrongly estimated, it will invalidate the

entire network topology. In addition, wrong topology may result from the simultaneous erroneous identification of several breakers. Hence, 10000 topology reconstructions were considered with 1 to 10 missing breaker status signals.

Table 5.1: Results from applying the Global autoencoder in reconstruction of 1-10 missing breaker statuses. In breaker status reconstruction, the stability of efficiency is not dependent on the number of missing signals. However, topology reconstruction significantly deteriorates with an increasing number of missing signals.

Nº missing signals	**Breaker status reconstruction**			**Topology reconstruction**		
	Wrong	Correct	Efficiency	Wrong	Correct	Efficiency
1	165	9835	98.35%	165	9835	98.35%
2	318	19682	98.41%	309	9691	96.91%
3	413	29587	98.62%	405	9595	95.95%
4	571	39429	98.57%	542	9458	94.58%
5	777	49223	98.45%	727	9273	92.73%
6	1007	58993	98.32%	925	9075	90.75%
7	1256	68744	98.21%	1118	8882	88.82%
8	1503	78497	98.12%	1330	8670	86.70%
9	1898	88102	97.89%	1652	8348	83.48%
10	2330	97670	97.67%	1955	8045	**80.45%**
Total	10238	539762	98.14%	9128	90872	**90.87%**

Table 5.1 summarizes the statistical results related to the correct recovery of individual breaker status and the entire system topology. The results are promising and show significant capabilities of autoencoders to cope with the reconstruction of unknown breaker status from the available power measurements. As expected, the efficiency of a breaker status reconstruction does not variate with the augmentation of the number of missing signals. This also proves that the test case is sufficiently large. Regarding the topology reconstruction, the efficiency significantly deteriorates with a number of missing signals. This is expected considering that all the 10 breakers need to be correct in order to consider a correctly reconstructed topology. Finally, it can be concluded that the concept seems promising mostly due to an average breaker reconstruction efficiency of 98.14%, or even more in regard to 80.45% of topology reconstruction efficiency without any information about the breaker statuses.

5.1.2 Benefits from the Penalty Term

Unlike the analog measurements, the breaker status signal is a digital variable, so only 2 integer values, e.g. -1 and 1, are acceptable as a solution. Therefore, the optimization result should be restricted to one of these two values in order to avoid the ambiguous solutions (around 0) to be

obtained by the model. For this purpose, one proposes a penalty term to be added to the objective function $\varphi(\varepsilon)$ from equation (4.1), in order to favor the result in a vicinity of the two possible integer values of the topology variable and penalize the real values in between. The new proposed objective function ψ is defined as the following:

$$\psi(\varepsilon) = \varphi(\varepsilon) + k\sum_{i=1}^{S}(1-x_i^2) \tag{5.1}$$

where:

- ε the vector of errors between the input and output of the autoencoder,
- S the number of switch status signals,
- x_i a real-valued signal output from the autoencoder associated to a breaker status signal i,
- k the penalty scaling coefficient,
- $\varphi(\varepsilon)$ the objective function from equation (4.1).

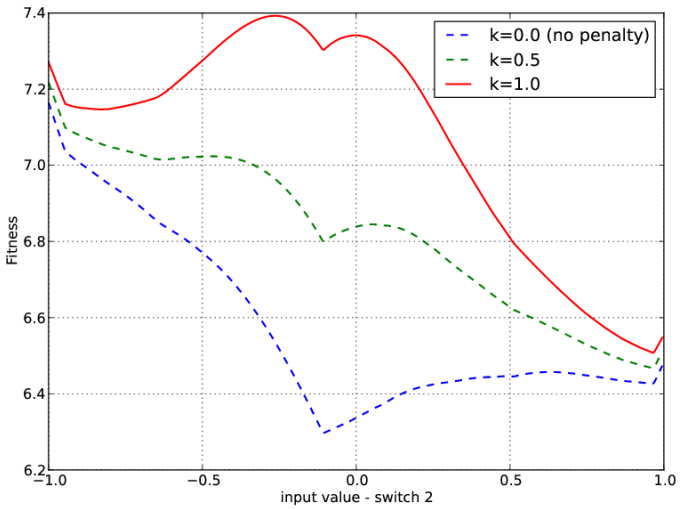


Figure 5.2: Example of the effect of the penalty coefficient k in the shape of the objective function to be minimized: a higher k raises the value in the center of the interval $[-1,1]$ and displaces the global minimum from this region to the correct value 1. Without penalty, the minimum is at a negative value and rounding leads to an incorrect answer.

Addition of the penalty term changes the shape of the objective function by pushing the minimum values as close as possible to -1 or 1. For instance, Figure 5.2 shows a representative objective function for a single sensor missing. In this figure, the x-axis represents the input variable x in the range $[-1,1]$, and the y axis represents value of the cost function $\psi(\varepsilon)$ in the last iteration, as a function of x (equation 5.1), for three values of penalty scaling coefficient $k = \{0, 0.5, 1\}$. In the represented case, it can be observed that the penalty had contributed to the reconstruction success. More specifically, an optimization to recover the missing signal, without the penalty term (coefficient $k = 0$ in equation 5.1) would give an optimum close to the center (blue dashed line) but slightly leaning to the left. Rounding this value to the closest extreme would give a reconstructed

signal of -1. With $k = 0.5$ and $k = 1$, one already has a definition of $x = 1$ as the optimum, which is the correct solution in this particular case.

In order to test the real benefits of the proposed penalty term, the same trained Global autoencoder is applied to the same test dataset with the penalty term included in the objective function. Results from Table 5.2 reveal notable benefits regarding both breaker and topology reconstruction efficiencies compared to results from Table 5.1. The difference is visible the most in the case of 10 missing breakers, where the penalty term had increased the topology reconstruction efficiency from 80.45% to 83.70%.

All the experiments have led to the conclusion that already a moderate coefficient k is enough to ensure good corrective effects, induced by the penalty term. Since augmenting the coefficient k until 1 does not deteriorate the efficiency, in this and all the future experiments the coefficient k is set to 1.

Table 5.2: Breaker status and topology reconstruction with the Global autoencoder, including the penalty term. The results prove significant benefits of the penalty term.

Nº missing signals	**Breaker status reconstruction**		**Topology reconstruction**	
	w/o Penalty	**w/ Penalty**	**w/o Penalty**	**w/ Penalty**
	Efficiency	Efficiency	Efficiency	Efficiency
1	98.35%	98.69%	98.35%	98.69%
2	98.41%	98.73%	96.91%	97.52%
3	98.62%	98.86%	95.95%	96.71%
4	98.57%	98.81%	94.58%	95.44%
5	98.45%	98.72%	92.73%	93.93%
6	98.32%	98.67%	90.75%	92.63%
7	98.21%	98.54%	88.82%	90.69%
8	98.12%	98.44%	86.70%	88.81%
9	97.89%	98.31%	83.48%	86.67%
10	**97.67%**	**98.09%**	**80.45%**	**83.70%**
Total	**98.14%**	**98.48%**	**90.87%**	**92.48%**

5.1.3 Global vs. Local Autoencoder

The proposed auto-associative topology estimator resides in the paradigm that an autoencoder is able to extract information about the system topology from the available electric values. If the entire system is observed, most of the available measurements are relevant for topology identification. However, some burdens of such a global model need to be noted. A number of weights that need to be trained in an autoencoder will grow exponentially with the size of the

power system. Also, the number of breakers, or possible topology variants may pose a significant challenge in the topology estimation process.

Therefore, after the preliminary method proposal, the intention is to eliminate the identified limitations by decentralization of the method, motivated by the following issues:

- ▷ impact of a breaker status change on the power and voltage measurements is the strongest in the narrow local area and vanishes with increasing electrical distance from a breaker,
- ▷ power system topology manipulation is restricted to a local area, e.g. a single substation or a single feeder.

In regard to these issues, information about breaker status has relatively strong local properties, i.e. local electric values are expected to be highly informative, while some remote measurements are expected to be less relevant. Moreover, in a particular case study, the status of each breaker is defined independently and the identification of topology as a whole does not facilitate the task, but rather broadens the dimensionality of the problem.

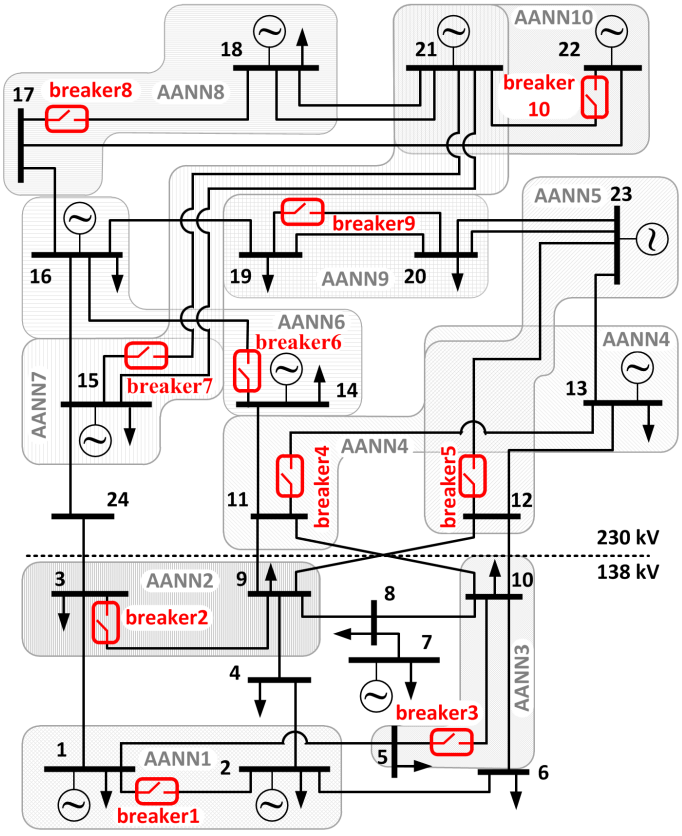


Figure 5.3: IEEE RTS 24-bus system with 10 breakers and indication of the areas of measurement collection used for training of 10 local autoencoders.

In accordance with these arguments, it can be speculated that an autoencoder based on local information should reach at least as good performances as the global autoencoder. A new term,

Local Autoencoder, relates to AANN which relates to topology of a sub-network considering a reduced set of data. For example, in a particular case study on the 24-bus system with 10 breakers, a local autoencoder would be in charge of a single breaker, while considering measurements from the local surrounding. In contrast, a global autoencoder uses the whole set of measurements available at the SCADA in order to perform breaker status recognition and, in theory, it is supposed to address the status of all breakers in the system. Therefore, instead of a single global autoencoder, multiple local autoencoders would be defined which should all together be able to reconstruct system topology in a decentralized way.

The described concept is tested in the same case study used in section 5.1.1. Since each breaker status is defined independently, 10 local self-tuning autoencoders are defined for the 10 breakers. Obviously, inputs for each autoencoder need to be chosen from measurements close to a particular breaker. Thus, only active/reactive power measurements from the lines and buses incident to a breaker are considered. This results in a mosaic of local autoencoders, as depicted in Figure 5.3.

The results of reconstruction from 1 to 10 missing signals with 10 local autoencoders (the penalty term included) are given in Table 5.3. If compared to Table 5.2, it becomes evident that the mosaics of local autoencoders exhibit considerable superiority over the global autoencoder approach. More efficient average signal reconstruction (98.48% to 99.48%) results in significant improvement in topology reconstruction, which is especially becomes evident with 10 missing signals when efficiency of 91.07% is achieved, compared to 83.7% of the global model.

Table 5.3: Results from applying 10 local self-tuning autoencoders in recomposition of 1-10 missing breaker status. The results are obtained with consideration of the penalty term. A significant improvement is visible when compared to the global model results from Tables 5.1 and 5.2.

Nº missing signals	**Signal reconstructions**			**Topology reconstructions**		
	Wrong	Correct	Efficiency	Wrong	Correct	Efficiency %
1	99	9901	99.01%	99	9901	99.01%
2	187	19813	99.06%	187	9813	98.13%
3	287	29713	99.04%	285	9715	97.15%
4	363	39637	99.09%	360	9640	96.40%
5	469	49531	99.06%	464	9536	95.36%
6	562	59438	99.06%	557	9443	94.43%
7	629	69371	99.10%	624	9376	93.76%
8	719	79281	99.10%	708	9292	92.92%
9	817	89183	99.09%	805	9195	91.95%
10	906	99094	**99.09%**	893	9107	**91.07%**
Total	5038	544962	**99.08%**	4982	95018	**95.02%**

Apart from accuracy, a significant gain is also obtained regarding the training requirements.

For example, the global autoencoder with architecture of 126-60-126 has 15120 weights to be trained in total. At the same time, the sizes of local autoencoders range from the minimum of 13-8-13 to a maximum of 19-14-19, which makes 3588 weights in total. Hence, around 4 times fewer weights need to be trained in the local model and all AANN can be trained in parallel. Furthermore, the local concept allows for more flexibility in the case of a frequently changing dynamic system. In regard to the constrained search optimization, in the case of multiple missing signals, a global autoencoder requires a very demanding multivariate search. Alternatively, in case of a local concept, this problem is split in multiple independent unknowns. This significantly accelerates the search and makes the result accurate. Finally, the global concept is more difficult to be applied on large scale systems due to exponentially rising training requirements. The concept of a mosaic of local autoencoders is however directly scalable to systems of any size.

In summary, the topology estimation problem obviously needs to be approached locally. Benefits that can be gained over a Global concept are resumed here one more time:

- ◇ Reduced training time, due to a reduction of weights to be trained,
- ◇ Reduced number of unknown variables which are being reconstructed by the unconstrained optimization,
- ◇ Improved missing signal efficiency, which is a result of the former two issues,
- ◇ Increased robustness to bad data,
- ◇ Increased flexibility of the model and scalability to large-scale systems.

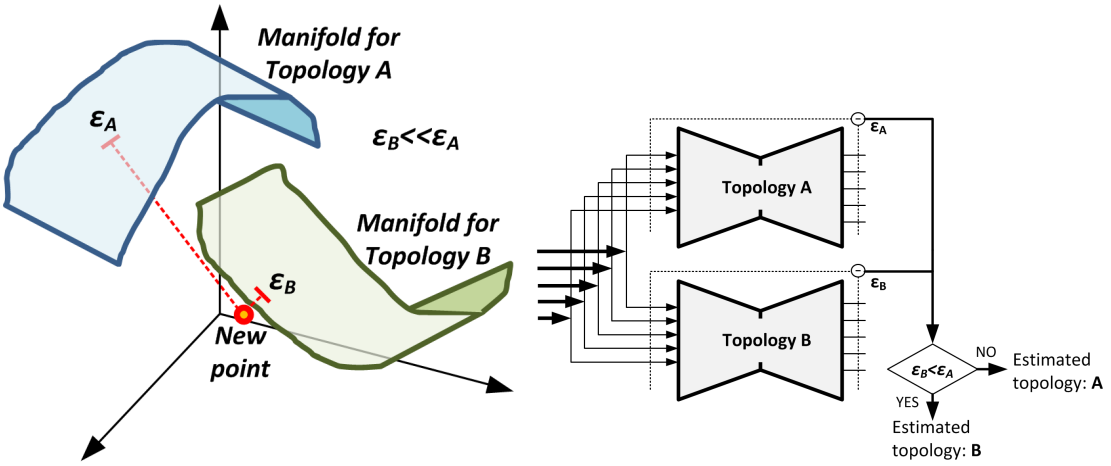
5.1.4 Competitive vs. Self-tuning Local Autoencoder Scheme

The proposed model of local self-tuning autoencoders has provided very promising results, and seems quite general and applicable to any system. However, a method proposed in [150] for transformer fault diagnosis was recognized as a possible alternative that may be even more suitable for the problem of topology reconstruction. The *Competitive auto-associative scheme*, is proposed in this thesis for application in power system topology estimation (as reported in [158]).

The competitive auto-associative scheme is used in [150] for recognition of transformer faults. In general, this method may be used for any pattern recognition problem in the following way: autoencoder is trained for each pattern with the corresponding features. Afterwards, in the real time, a new feature vector is presented to all trained autoencoders and the one which results in the least input-output error is proclaimed a winner, indicating thus a recognized pattern.

The main idea behind the novel concept is to perceive topology estimation as a pattern recognition problem, where each topology defines a manifold in the space of available electric variables. As illustrated in Figure 5.4(a), manifolds for topologies A and B are learned by the two corresponding autoencoders. When a new point is presented to the both, the one which results in a smaller input-output error indicates that the point pertains to its manifold. An autoencoder is labeled for a particular topology and is trained with measurements which correspond to that

topology. The recognition stage relies on the fact that only a specific autoencoder will resonate with an unclassified sample, while others will show a significant auto-association error (Figure 5.4(b)). In other words, when a pattern of relevant measured data is shown to all trained autoencoders, one of them will respond with a very small error (the input vector will be in tune with one of the patterns learned by the autoencoders), thus indicating the unknown topology.



(a) Manifolds of two distinctive topologies. (b) Competitive autoencoders for topology estimation.

Figure 5.4: Illustration of a competitive auto-associative topology estimation scheme: (a) Two manifolds defined by topologies A and B, learned by autoencoders. The autoencoder whose manifold is closer to a new point will respond with substantially smaller input-output error then the autoencoder which identifies this input as an outlier; (b) Scheme for the competitive auto-associative topology estimator.

In order to evaluate the novel concept, it was tested using the same case study of a 24-bus system with 10 breakers from Figure 5.1. If a global model is used, 1024 competitive global autoencoders for 1024 topologies would be required. A model of this type is obviously not sustainable in large scale, so it is decentralized and localized the same as in the previous section. Since each breaker can take 2 possible states (topologies), it is associated with two autoencoders: one trained with local data when the status is closed, and the other trained for the status open. Afterwards, when a relevant measurements vector is presented to the autoencoders, the one with lower input-output errors should indicate the correct breaker status. Boarders for the input measurements selection are the same as shown in Figure 5.3, i.e. the same as in the self-tuning method. Unlike the self-tuning concept, there are now 2 autoencoders trained per each breaker, and the number of inputs is decreased by 1, since there is no need for a breaker status signal on the input.

The results presented in Table 5.4 reveal a substantial efficiency improvement when compared to the self-tuning model. With an average breaker status reconstruction efficiency of 99.97%, the topology is estimated correctly in 99.72% of cases without any prior knowledge of breaker statuses. This is an impressive improvement when compared to 91.07% of the self-tuning model.

Moreover, it is of key importance to note that a competitive scheme provides an enormous gain regarding the computational efficiency. For example, reconstruction of the competitive scheme is practically instantaneous, since the computational time necessary to obtain an output of a neural network is practically negligible. On the other hand, the self-tuning model requires an optimization method which may pose an insuperable burden for a state estimation pre-processing applicability.

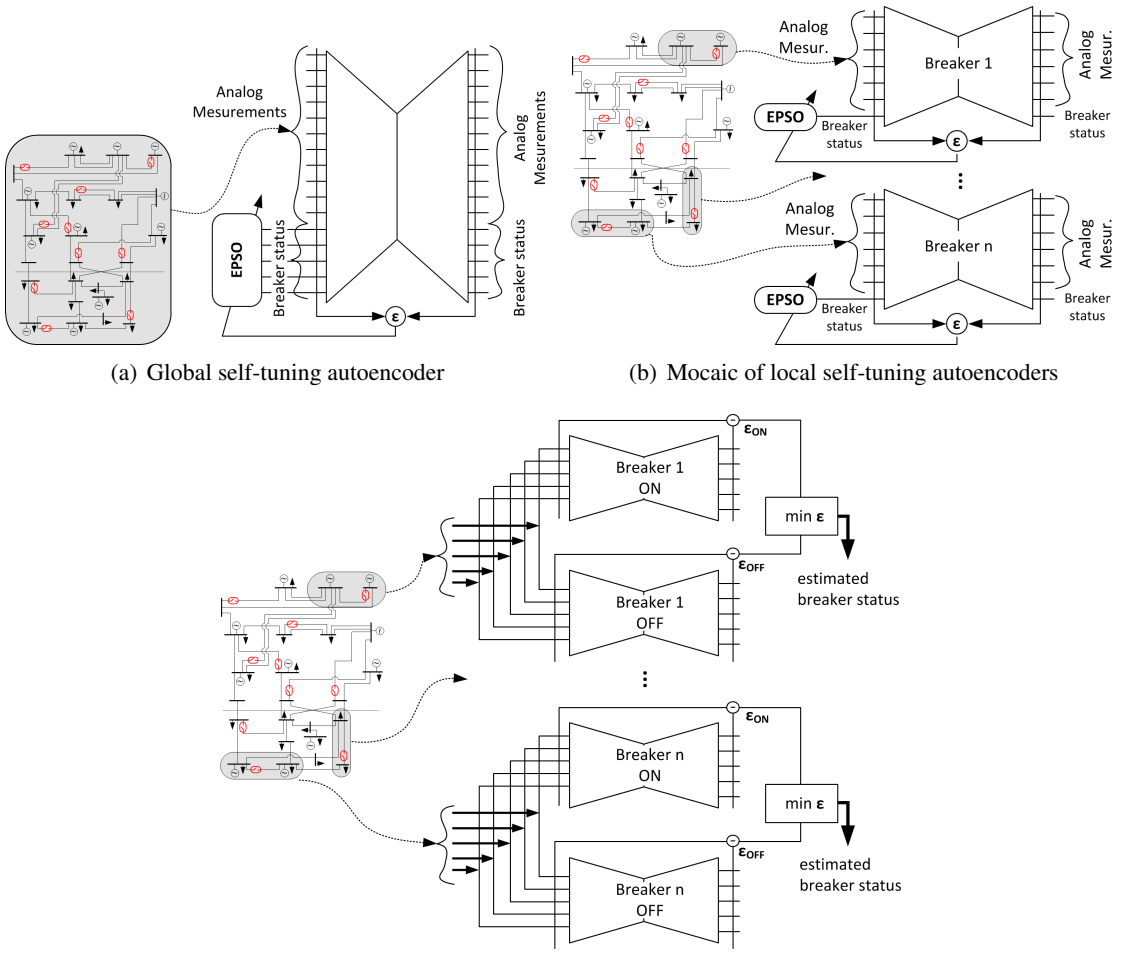
Table 5.4: Results from application of 10 local competitive autoencoders in recomposition of 1-10 missing breaker status. Enormous efficiency enhancement is achieved comparing to the self-tuning model from Table 5.3.

Nº missing signals	**Signal reconstructions**			**Topology reconstructions**		
	Wrong	Correct	Efficiency	Wrong	Correct	Efficiency
1	3	9997	99.97%	3	9997	99.97%
2	5	19995	99.98%	5	9995	99.95%
3	8	29992	99.97%	8	9992	99.92%
4	11	39989	99.97%	11	9989	99.89%
5	14	49986	99.97%	14	9986	99.86%
6	15	59985	99.98%	15	9985	99.85%
7	18	69982	99.97%	18	9982	99.82%
8	24	79976	99.97%	24	9976	99.76%
9	26	89974	99.97%	26	9974	99.74%
10	28	99972	**99.97%**	28	9972	**99.72%**
Total	152	549848	**99.97%**	152	99848	**99.85%**

The three analyzed auto-associative concepts: Global self-tuning, Local self-tuning and Local Competitive are illustrated in Figure 5.5. After the testing of methods in topology reconstruction of the 24-bus system with 10 breakers the following can be summarized:

- ◇ The results of the most challenging topology estimation efficiency with 10 unknown breakers: 83.70%, 91.07% and 99.72% for global self-tuning, local self-tuning and competitive models, respectively.
- ◇ Global model is not only less accurate when compared to local, but also much more computationally demanding and thus barely applicable in large scale systems.
- ◇ When compared to the self-tuning model, the proposed competitive scheme offers a significant improvement in efficiency without any trade-offs.
- ◇ Competitive model requires more autoencoders to be trained, but unlike the self-tuning model, the topology estimation process is practically instantaneous and independent of the system size.

Based on the results from preliminary tests, it can be concluded that the competitive auto-associative model seems most promising for application in the real topology estimation procedure. The derived conclusions are verified hereafter in more demanding and more realistic case studies.



(a) Global self-tuning autoencoder

(b) Mocaic of local self-tuning autoencoders

(c) Mocaic of local competititve autoencoders

Figure 5.5: Illustration of auto-associative topology estimation concepts: (a) Global self-tuning concept; (b) Mosaic local self-tuning concept; (c) Mosaic local competitive concept.

5.2 Competitive Auto-associative Mosaic - Method Verification

In the former section, the competitive auto-associative mosaic method is shown to be superior to self-tuning regarding both topology estimation and the computational efficiency. This section performs more exhaustive tests in order to verify the method and justify its contributions to power system topology estimation.

5.2.1 Identification of a Single Breaker Status - High Redundancy

In order to investigate the results from the former section in more detail, the performance of the method is tested on reconstruction of each particular breaker status. Table 5.5 presents the status identification results for each of the 10 breaker locations on a sample of 10000 distinct scenarios. The results confirm that a competitive scheme performs superior to all 10 breakers. The table also reveals architectures of each autoencoder in the self-tuning model. The competitive autoencoders have one input/output less then the self-tuning ones, because the breaker status signal does not need to be included.

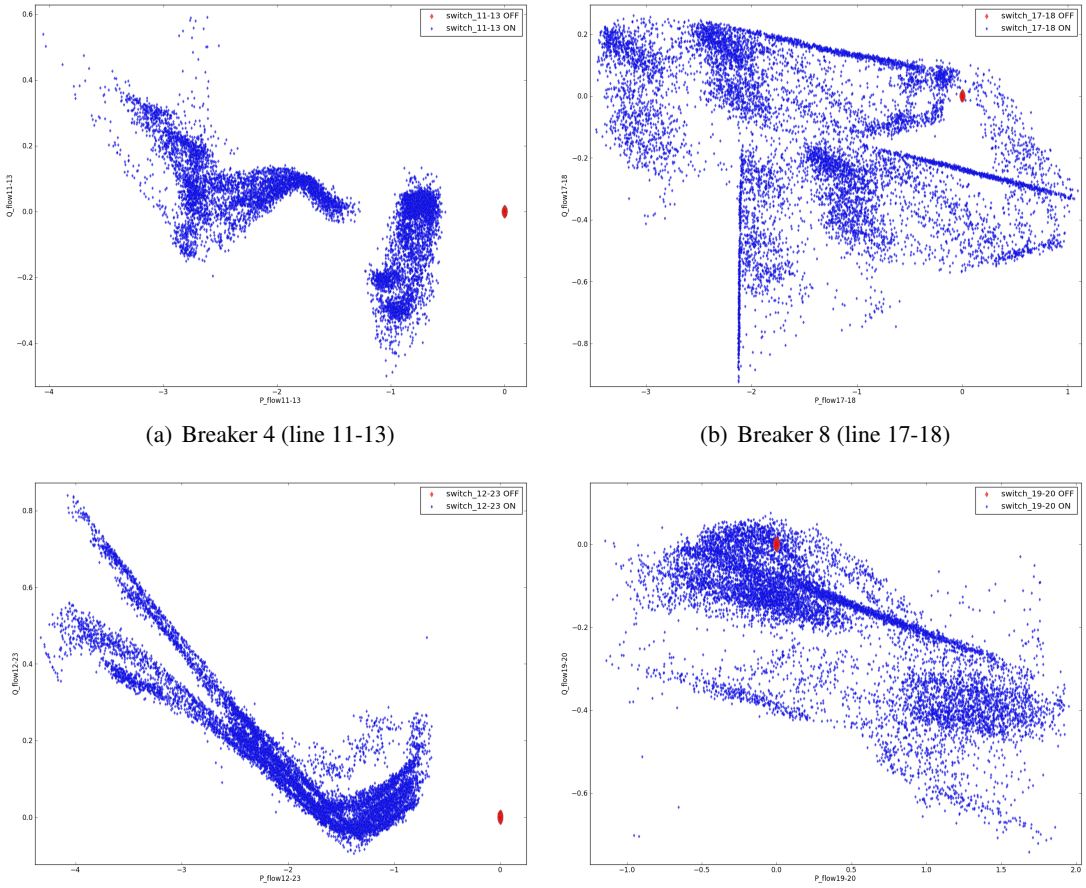
Table 5.5: Comparison of the self-tuning and competitive models for topology identification with local autoencoders.

Breaker	**Self-tuning**					**Competitive**		
	Input/ Output	Hidden layer	Wrong	Correct	Efficiency	Wrong	Correct	Efficiency
1	15	10	0	10000	100.00%	0	10000	100.00%
2	19	14	1	9999	99.99%	0	10000	100.00%
3	17	12	162	9838	98.38%	17	9983	99.83%
4	15	12	0	10000	100.00%	0	10000	100.00%
5	15	12	0	10000	100.00%	0	10000	100.00%
6	15	10	0	10000	100.00%	0	10000	100.00%
7	17	12	0	10000	100.00%	0	10000	100.00%
8	11	8	381	9619	96.19%	7	9993	99.93%
9	13	8	277	9723	97.23%	0	10000	100.00%
10	17	12	0	10000	100.00%	0	10000	100.00%
Total			821	99179	**99.18%**	24	99976	**99.98%**

It may also be observed that the degree of difficulty for the breaker status disclosure from the measured analog data is not the same for all of the breakers. The reasons behind this fact still deserve further investigation, but the problem is likely to be related to the selection of a measurement set selected for inputs of each autoencoder. An optimization procedure for composing such a set of most relevant measurements would undoubtedly be an useful method upgrade.

In regard to the estimation imperfection of some autoencoders, one needs to consider that the cause of failure may be also in the ambiguity of the breaker status cases. For example, it is possible that active/reactive power flow in the breaker line approximates to zero, although the breaker is closed. In these cases, breaker status is barely identifiable, since power flows in all the system almost do not change at all by opening/closing the specific breaker. In order to examine the

possibility of such cases to occur, a scatter plot in Figure 5.6 shows active/reactive power flows measured through the line where a particular breaker is located for breakers 4, 5, 8 and 9. Points for closed and open breaker statuses are blue and red, respectively. It can be observed that in cases of breakers 4 and 5, open and closed breaker statuses are clearly separated and status identification seems trivial. Trivial means that even a heuristic method should not have difficulty in discovering a breaker status only by knowing the active/reactive power flows in the particular breaker lines.



(a) Breaker 4 (line 11-13)

(b) Breaker 8 (line 17-18)

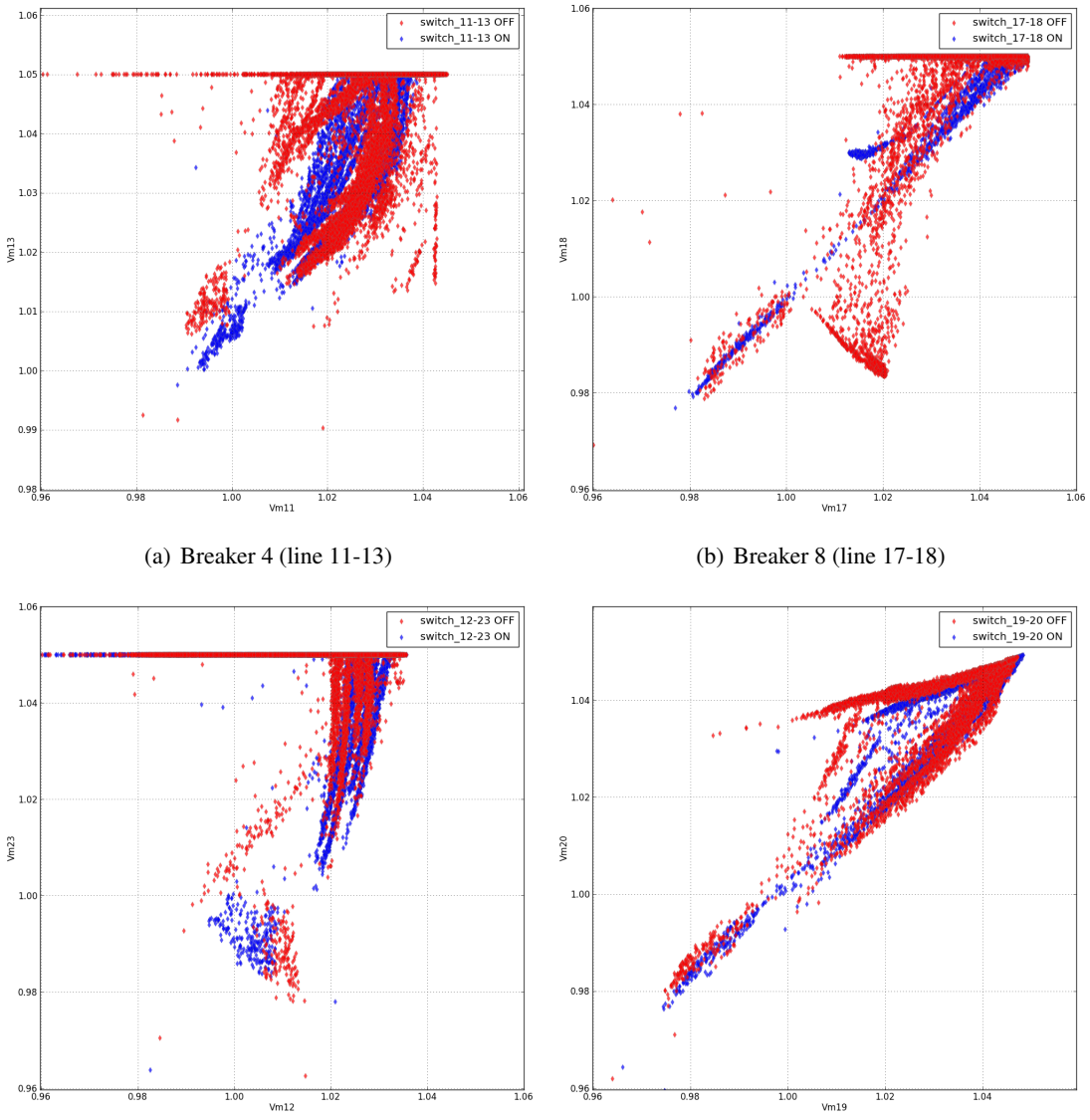
(c) Breaker 5 (line 12-23)

(d) Breaker 9 (line 19-20)

Figure 5.6: Scatter plot of active-reactive power flow in the breaker line for ON and OFF status of breakers 4, 5, 8 and 9. It is important to observe an overlap between ON and OFF scenarios of breakers 8 and 9, while in the case of breakers 4 and 5 scenarios are easily distinguishable.

On the other hand, for the breakers 8 and 9 (Figures 5.6(b) and 5.6(d)) a certain overlap between the scenarios with an open or closed breaker can be noted. This means that cases when the breaker is closed and the power flow through the relevant line approaches zero, are probable to occur. However, if the breaker status is open, the power flow through its line is obviously zero but it does not mean that other electrical variables correspond to the closed breaker status with a zero flow. In other words, the power flow through the breaker line is not the only relevant information and other electrical values could also contribute to breaker status identification. For example, if a

breaker is closed, and the flow through it is zero, voltage magnitudes on the incident buses need to be equal. On the other hand, open breaker status does not guarantee equality of voltage magnitudes on the incident buses. For the same four breakers, a scatter plot of voltage magnitudes on incident nodes for open and closed (OFF and ON) breaker status is presented in Figure 5.7. Information content in voltage magnitudes about breaker status is evident.



(a) Breaker 4 (line 11-13)

(b) Breaker 8 (line 17-18)

(c) Breaker 5 (line 12-23)

(d) Breaker 9 (line 19-20)

Figure 5.7: Scatter plot of voltage magnitudes incident to breakers for ON and OFF breaker status for breakers 4, 5, 8 and 9. The purpose of the plot is to visualize the information about a breaker status which is hidden in the incident voltage magnitudes.

If a breaker line power flow is known, it seems trivial to identify a breaker status. This hypothesis is tested in the following case study by comparing the competitive autoencoders to a heuristic method. More specifically, the heuristic method represents quite a primitive way of

breaker status identification according to breaker power flow only. It defines boundary limits for the active/reactive power flow when the breaker is closed. If active and reactive power flows of the breaker line are within or outside the limits, the breaker is identified as closed or open, respectively. On the other hand, besides the power flow through the breaker, the autoencoder considers all incident active/reactive injection and power flows plus voltage magnitudes on the buses incident to a breaker. The autoencoder is expected to perform at least as good as the heuristic method. The comparative results are shown in Table 5.6. The case study is the same as for the results from Table 5.5, except that the competitive additionally considers voltage magnitude measurements.

It is important to note that knowledge about the power flow through the breaker allows the application of a quite primitive heuristic method which is able to perform almost as efficiently as the autoencoders. It can also be noted that the heuristic method does not provide an upper boundary for breaker reconstruction efficiency. For example, in case of the breaker 9, heuristic method is confused in 18 cases, while competitive autoencoders always manage to correctly identify the breaker status. This indicates that more than a power flow through breaker is needed for a breaker status reconstruction. In other words, the rest of the measurements are not just redundant with the breaker power flow but can provide some additional information. For example, open breaker status will mostly result in some discrepancy in incident voltage magnitudes, while closed breaker status with a zero flow implies equivalent voltages.

Table 5.6: A primitive heuristic method against the competitive autoencoders in breaker status reconstruction. If power flows through the breakers are available, a heuristic method is able to perform as well as the autoencoders.

Breaker	**Competitive AANNs**			**Heuristic**		
	Wrong	Correct	Efficiency	Wrong	Correct	Efficiency
1	0	10000	100.00%	0	10000	100.00%
2	0	10000	100.00%	0	10000	100.00%
3	3	9997	99.97%	2	9998	99.98%
4	0	10000	100.00%	0	10000	100.00%
5	0	10000	100.00%	0	10000	100.00%
6	0	10000	100.00%	0	10000	100.00%
7	0	10000	100.00%	0	10000	100.00%
8	16	9984	99.84%	1	9999	99.99%
9	0	10000	100.00%	18	9982	99.82%
10	0	10000	100.00%	0	10000	100.00%
Total	19	99981	**99.98%**	21	99979	**99.98%**

Another issue of key importance is that the heuristic method is applicable only in cases when power flow through the breaker is available, otherwise, a more sophisticated method should

be used. Some of the classical methods, based on mathematical programming are expected to solve such a problem, but only if system is observable in the breaker surrounding. On the other hand, autoencoders do not require observability in a classical sense since they only extract the information about topology through available electric measurements without requiring a mathematical model of a system. Therefore, the proposed method is expected to be especially contributive in cases of full observability lack or poor measurements redundancy. This is intended to be demonstrated in the subsequent case studies.

5.2.2 Identification of a Single Breaker Status - Flows Through Breakers are not Available

Results discussed in the former section seem quite concerning, given that a primitive heuristic method is capable to estimate breaker status almost as efficiently as the proposed method. This is not strange, since breaker status reconstruction seems quite trivial if active/reactive power flows through the breaker are assumed to be available. More specifically, the power flow through the breaker line mostly contains a direct information about the breaker status. This information is a prerequisite for a heuristic method, and the intention of the test case presented here is to demonstrate that this is not a prerequisite for autoencoders. For that purpose, it is assumed that active/reactive power flows through all the 10 breakers are simply not available.

Table 5.7: Results of breaker status estimation with competitive autoencoders without availability of breaker power flows. If compared to results from Table 5.6, the deterioration of accuracy is insignificant despite the exception of the direct information. Also, if voltage magnitudes from the incident nodes are available, a slight improvement is achieved.

Breaker	**w/o $\lvert V_i\rvert$ and $\lvert V_j\rvert$**			**w/ $\lvert V_i\rvert$ and $\lvert V_j\rvert$**		
	Wrong	Correct	Efficiency	Wrong	Correct	Efficiency
1	0	10000	100.00%	0	10000	100.00%
2	76	9924	99.24%	31	9969	99.69%
3	33	9967	99.67%	7	9993	99.93%
4	0	10000	100.00%	0	10000	100.00%
5	0	10000	100.00%	0	10000	100.00%
6	0	10000	100.00%	0	10000	100.00%
7	0	10000	100.00%	0	10000	100.00%
8	34	9966	99.66%	30	9970	99.70%
9	0	10000	100.00%	0	10000	100.00%
10	0	10000	100.00%	0	10000	100.00%
Total	143	99857	**99.86%**	68	99932	**99.93%**

Table 5.7 provides results of a particular breaker status reconstruction for two measurement sets: one considers only incident power measurements; and the other additionally includes incident voltage measurements ($|V_i|$ and $|V_j|$ for breaker in the line $i-j$). It can be noted that, despite the absence of key information, only a slight degradation in the result for the breakers 2, 3 and 8 can be observed, regarding the results from Tables 5.6 and 5.5. For the other breakers, impeccable efficiency is maintained. Also, as expected, the consideration of voltage magnitude information has slightly contributed to efficiency enhancement in the critical cases. Results of this experiment imply two significant conclusions:

- Autoencoders manage to capture all relevant information about the breaker status in their weights, and the competitive scheme provides a robust breaker status reconstruction.
- Consideration of voltage magnitude measurements reveals a sensor fusion capacity of an autoencoder. In other words, the proposed auto-associative topology estimation could be applied to any kind of measurements which provide information about topology, i.e. it can be even used with PMU measurements.

5.2.3 Identification of a Single Breaker Status - Remote Measurements Only

An additional experiment is performed in order to test performance of competitive autoencoder in an extreme case of low local measurement redundancy. Here, all the most relevant, adjacent local measurements, selected for inputs of local autoencoders in the former examples (Figure 5.3) are assumed in this study to be unavailable. Instead, only remote power flow measurements are employed, from the second order of adjacency of each breaker. For example, Figure 5.8 indicates 14 active and reactive power flow measurements considered for inputs of the 2 autoencoders in charge of the breaker number 6. The measurement availability for the other 9 breakers is determined in the same manner.

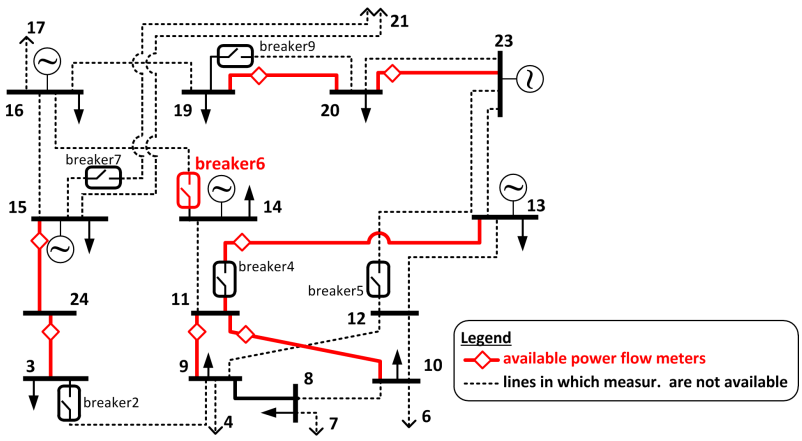


Figure 5.8: Partial representation of the IEEE RTS 24 system, with indication of measurements selection for reconstruction of breaker 6. Flow measurements are available only for solid, red lines and not for dashed lines — only remote measurements are considered.

It is of key importance to mention that in this case study the lack of observability in the breaker surrounding was provoked. In other words, each particular breaker resides in the middle of an unobservable island. Therefore, any of the post-processing topology estimators, or any method based on mathematical system modeling is not able to work in this case.

Table 5.8: Reconstruction of 10 breaker statuses, considering only remote measurements. Despite the fact that each breaker is in the middle of unobservable island, the method is very efficient for most of the breakers. Superiority of competitive model over self-tuning is once again demonstrated.

Breaker	**Self-tuning**					**Competitive**		
	Input/ Output	Hidden layer	Wrong	Correct	Efficiency	Wrong	Correct	Efficiency
1	13	10	3536	6464	64.64%	2232	7768	77.68%
2	19	15	1479	8521	85.21%	1051	8949	89.49%
3	21	17	1837	8163	81.63%	1235	8765	87.65%
4	19	15	1604	8396	83.96%	286	9714	97.14%
5	15	12	3191	6809	68.09%	2504	7496	74.96%
6	**15**	**12**	**19**	**9981**	**99.81%**	**0**	**10000**	**100.00%**
7	13	10	3244	6756	67.56%	2697	7303	73.03%
8	11	9	1474	8526	85.26%	468	9532	95.32%
9	13	10	4668	5332	53.32%	4449	5551	55.51%
10	11	9	168	9832	98.32%	119	9881	98.81%
	Total		21220	78780	78.78%	15041	84959	84.96%

The results in Table 5.8 demonstrate breaker reconstruction efficiency when considering only remote measurements. Once again, superiority of the competitive scheme against the self-tuning model is justified. What is much more indicative is that even when the most relevant information is unavailable, the method is able to achieve a very high efficiency. Moreover, it is important to understand that any classical method that relies on mathematical model of the power system is not able to work in this case due to lack of observability.

The method efficiency significantly variates from breaker to breaker, thus raising again concerns about the issue of detecting the most informative input measurements. Namely, information about a breaker status is spread throughout the analog measurements, depending on the system parameters and topology. Breakers with high reconstruction efficiency (breaker 6) are expected to have high spread of information about the status, while poor efficiency in case of other breakers indicate that the information is very narrowly concentrated to the local measurements. These results have raised an interest in development of a method for defining the most relevant input measurements selection regarding the topology variable of interest.

From the results obtained in this case study, the following conclusions can be derived:

- The proposed competitive auto-associative topology estimator is very robust in regard to the available measurement redundancy,

- Unlike the conventional topology estimators, the proposed model does not require system observability as a prerequisite, so it can be very useful in weakly monitored systems (e.g. distribution networks).

- A breaker status reconstruction efficiency will greatly impact the spread of information about topology through the available power (and/or voltage) measurements.

- Incidence is not a unique criterion for relevance of a measurement for breaker status reconstruction. A method should be developed to provide a measurement set which is the most informative about the breaker status.

- Finally, it is now safe to say that the competitive scheme is far superior to the self-tuning model.

5.2.4 Robustness of the Competitive Model to Gross Errors

Performance of the proposed auto-associative topology estimator method is additionally verified by submitting it to a stress test of robustness to bad data. More specifically, in each of 10000 test case scenarios, each local autoencoder is contaminated with 1 to 5 randomly generated gross errors, uniformly distributed in $[-0.2,-0.1]\cup[0.1,0.2]p.u.$ As expected, results from Table 5.9 demonstrate a deterioration of estimation efficiency with an increase in number of GE. In addition, robustness obviously significantly varies from breaker to breaker. For example, some of the autoencoders preserve a very high performance and remain almost immune to 5 gross errors, which is an impressive result since 5 GE make almost 30% of the inputs. On the other hand, breaker number 8 for instance appears to be very sensitive to gross errors and collapses already in presence of 5 gross errors in the input.

It can be observed that the performed test is quite exaggerated since the probability of 5 GE to occur in a narrow local region is quite low probable. It would be more realistic to observe gross error occurrence in the system as a whole. In that case, the local models, as the proposed one, are in general more robust then the global ones.

Table 5.9: Analysis of gross errors impact on the efficiency of breaker status identification. Inputs of each autoencoder are corrupted with 1-5 gross errors. The results reveal very high robustness, and some autoencoders maintain the same efficiency even when almost 30% of inputs are corrupted with GE.

Breaker	Number of GE					
	0	1	2	3	4	5
	Breaker reconstruction efficiency					
1	100.00%	99.17%	97.31%	94.89%	92.98%	90.72%
2	100.00%	93.60%	89.40%	86.30%	84.23%	82.42%
3	99.83%	89.83%	82.67%	78.48%	75.49%	74.44%
4	100.00%	99.89%	99.58%	99.35%	99.01%	98.51%
5	100.00%	100.00%	100.00%	99.97%	99.95%	99.89%
6	100.00%	100.00%	100.00%	100.00%	100.00%	99.99%
7	100.00%	99.93%	99.84%	99.78%	99.56%	99.45%
8	99.93%	78.47%	67.35%	62.86%	59.71%	57.60%
9	100.00%	95.76%	92.10%	89.30%	86.66%	84.12%
10	100.00%	100.00%	99.99%	99.96%	99.91%	99.90%
Total	**99.94%**	**95.67%**	**92.82%**	**91.09%**	**89.75%**	**88.70%**

Section 4.2.3 proposed a concept of ***Denoising Auto-Associative Neural Networks*** (**DAANN**), which have been proven to be very robust and efficient in filtration of bad data. Unlike the normal autoencoder, a denoising autoencoder is not trained with the same input and target vectors. Instead, the input is corrupted with gross errors, while target vectors remain free of GE. With this type of training, DAANN learns not only the manifold, but also how to re-project the corrupted input onto the manifold. Such performance is expected to also contribute to robustness of the competitive topology estimation scheme.

In order to test the hypothesis, 20 competitive denoising autoencoders are trained for the same case study and tested on the dataset corrupted by 1 to 5 GE. Results are presented in Table 5.10 and the summarized comparative results are presented in Figure 5.9. It is evident that a significant improvement in robustness is gained by the denoising way of autoencoders training. For the corruption with 5 GE, the average efficiency of DAANN is 94.38% in comparison to 88.70% of the normal AANN. The advancement becomes even more evident when the entire system topology is observed. It is significant to observe the results for the breaker 8, where in the presence of 5 GE a normal autoencoder remains practically unusable (57.60%) while the denoising autoencoder achieves the efficiency of even 91.49%. Also, it is important to note that the denoising training did not lose the performances in the scenarios free of GE.

Table 5.10: Impact of GE on the efficiency of breaker status identification with competitive denoising autoencoders. It is important to note a substantial improvement in comparison to the performance of normal autoencoders from Table 5.9.

Breaker	Number of GE					
	0	1	2	3	4	5
	Breaker reconstruction efficiency					
1	100.00%	99.41%	97.73%	96.17%	94.17%	92.53%
2	99.94%	97.96%	95.47%	92.82%	91.05%	89.14%
3	99.58%	96.65%	94.16%	90.62%	88.95%	86.75%
4	100.00%	99.98%	99.95%	99.88%	99.78%	99.67%
5	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
6	100.00%	100.00%	100.00%	100.00%	99.99%	100.00%
7	100.00%	99.97%	99.91%	99.80%	99.71%	99.58%
8	99.82%	97.77%	95.90%	93.72%	92.42%	91.49%
9	99.76%	95.69%	92.38%	88.45%	86.44%	84.61%
10	100.00%	100.00%	99.99%	100.00%	99.99%	99.98%
Total	**99.91%**	**98.74%**	**97.55%**	**96.15%**	**95.25%**	**94.38%**

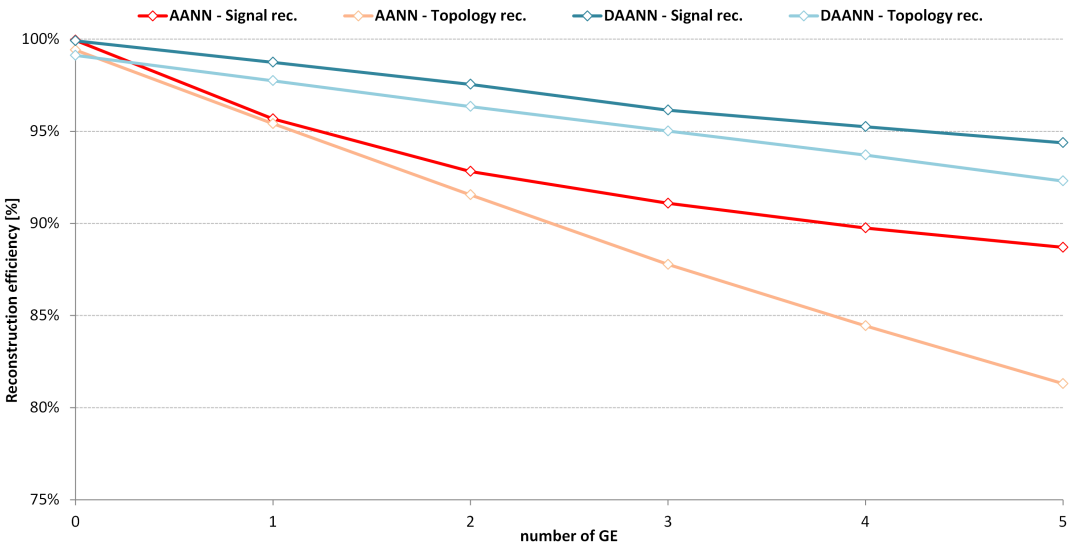


Figure 5.9: Gross error impact on the competitive autoencoder efficiency in signal reconstruction and system topology estimation. If the concept of denoising autoencoders is adopted, robustness is significantly improved, especially in the case of topology estimation. With 5 GE, the 10 DAANN couples manage to correctly estimate the topology in 92.3% of cases.

5.2.5 Proof of the Method Generality

The proposed method is tested on a variety of different considerations that can affect the topology estimation efficiency. However, all the previous tests were performed using the same dataset of measurements generated by the optimal load flow calculation and considering a predefined load distribution (as given in [130]). Accordingly, the following questions have been raised:

– *Is an autoencoder able to provide diligent results for any possible scenario in a power system?*

– *How is the proposed method expected to respond in some unexpected scenario like outage of a line, load or generator?*

– *What if power dispatch strategy is changed, or a new cable, load or generator is connected to the grid?*

All these questions are quite reasonable for any real, dynamic power system. Due to former tests, an autoencoder is capable to learn the data manifold, but what if a measurement vector scenario is far from the manifold and does not contain any bad data. This type of a vector would not be coherent with the autoencoder and the result of topology estimation becomes uncertain. For example, autoencoders are good in interpolation but if an outlier vector is presented to the input, they will respond with a significant input-output error, similar to the case of gross error presence. If none of the competitive autoencoders is coherent with the provided measurements vector, the reconstruction results will obviously be unsuccessful.

The case study performed here aims to prove that an autoencoder can be trained in such a way to be able to handle any scenario probable to occur in a real power system. In that regard, an autoencoder is intended to be trained with a comprehensive dataset, and then tested with data obtained from a specific load and generator scenario. The training dataset is generated with an aim to cover the most possible generation-load scenarios, i.e. the following was assumed:

- Each load was independently set from 30% to 100% with a uniform distribution,
- Active power production of each generator is randomly and independently set to a value in the range of minimum to maximum production with a uniform distribution,
- Certain number of cases include the outage of a randomly chosen generator and/or load,
- Reactive power production of each generator is independently and randomly set to a value in range of minimum-maximum production with a uniform distribution,
- Certain number of cases include an outage of randomly chosen generator and load,
- Statuses of the 10 breakers are defined randomly in each scenario and independently for each particular breaker.

When autoencoders are trained with random measurement scenarios, they are expected to learn at least the Kirchoff laws, in addition to some specific characteristics of a system. Afterwards, a test dataset is the same one used in all the former tests, generated using the optimal power flow and the specific load curve from [130]. Results of 10 breaker status reconstructions, considering full redundancy, are shown in Table 5.11. Obtained results prove the applicability of the method without concerns about specific events which may appear as outliers and harm the proposed competitive auto-associative topology estimator. In other words, it is shown that, if trained properly, the autoencoders are able to efficiently estimate breaker status independently of the scenario that may occur in the system.

Any other structural change in a power system can also be easily considered by the proposed method. Namely due to local decentralized model, a structural change will impact only a few of the autoencoders which can than be extended, restructured and re-trained. Finally, the concept is assessed as very flexible and easily scalable to large-scale systems. In the mean time, it is very robust in all barely predictable events like gross errors or unscheduled generation-load scenarios.

Table 5.11: Proof of autoencoder's generosity - test of autoencoders trained with general dataset which covers most of the probable load flow scenarios. The results indicate that the autoencoders are able to estimate breaker status with sufficient accuracy in any generation-load scenario.

Breaker	**Competitive AANN**		
	Wrong	Correct	Efficiency
1	0	10000	100.00%
2	111	9889	98.89%
3	188	9812	98.12%
4	0	10000	100.00%
5	0	10000	100.00%
6	0	10000	100.00%
7	0	10000	100.00%
8	26	9974	99.74%
9	104	9896	98.96%
10	0	10000	100.00%
Total	429	99571	**99.57%**

5.3 Topology Estimation in Substations - Split Bus Problem

In the former section, a topology estimation method is tested on quite a rudimentary sample. Topological variability of the system is simplified, being reduced to 10 breakers positioned in 10 power lines. In order to consider a more realistic case, a step back needs to be observed.

In fact, topology of an electric power system is defined by the status of circuit breakers within substations. Each substation may have different breaker configurations like: single-bus, double-bus/double-breaker, breaker-and-a-half, etc. After all breaker statuses are known, one can build a simplified, bus/branch model in which each substation is represented by the smallest possible number of buses. Such a simplified model is more similar to the system model used so far with fixed number of buses and 10 switches (Figure 5.1).

This section intends to apply the proposed auto-associative topology estimation method to the level of complex breaker arrangement within substations. This problem is addressed hereafter as the *Split bus* problem. Therefore, instead of a single breaker status reconstruction, observed in section 5.1, one will zoom into a substation and try to discover its topology. Therefore, the abilities of this type of a topology estimation method are directly conditioned with location of power flow meters. In fact, there are two possible measurement schemes:

- **External** - power flow meters are located outside of the substation, measuring the external power flows and internal measurements are not available,
- **Internal** - each circuit breaker is associated to a power flow meter, so both internal and external measurements are available.

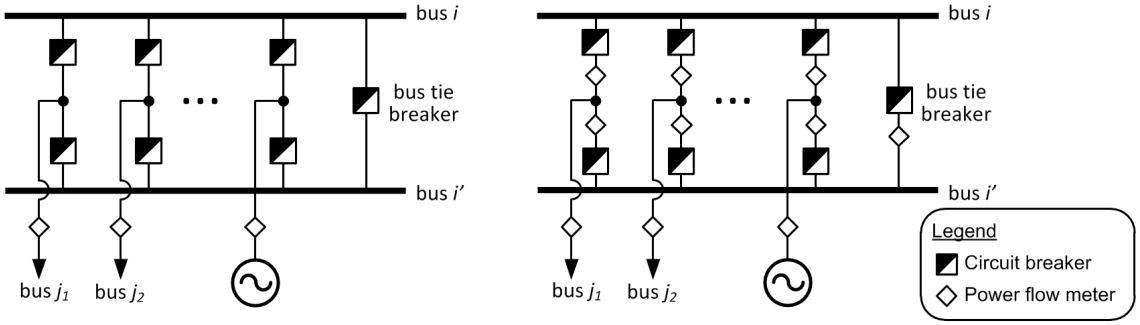


Figure 5.10: Two measurements schemes for a double-bus/double-breaker breaker arrangement with a bus tie breaker at a substation *i*. Left: external measurements - power flow meters are located out of the substation; Right: internal measurements - each circuit breaker has an associated power flow meter.

Figure 5.10 illustrates a substation topology with a double-bus/double-breaker scheme with both external and internal power flow meters scheme. In this regard, the most important issue is that it is impossible to identify a status of each particular breaker only with external measurements. It is only possible to determine topology of a substation which is compatible to the available external measurements. On the other hand, if one wants to deal with breakers, an additional information provided by the internal meters is needed.

In practice, internal power flow measurements are common, but the information which is generally transmitted to the state estimator comes mostly from the outside of the substation. Once again, it is important to emphasize that internal power flow information allows the identification of

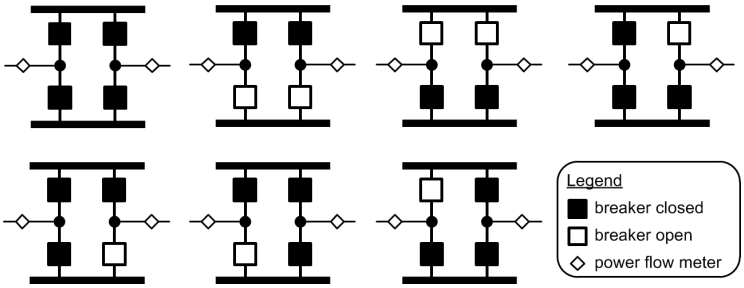


Figure 5.11: Configurations of a substation with 4 circuit breakers which are indistinguishable from the external power measurements.

the statuses of all breakers within each substation. On the other hand, the external measurements cannot identify a specific topology but only a group of them which results in the same external electric values. For the purpose of clarity, Figure 5.11 shows 7 different configurations of a substation with 4 circuit breakers, which all result in identical external measurements. Hence, these 7 configurations are indistinguishable if only external measurements are known.

The research work proposed in this section is also reported in [154] and [158]. The breaker status reconstruction problem, analyzed in section 5.1 can be seen as a proof of concept, and this section accommodates the same approach to a more realistic split bus problem. The method will be tested for both internal and external measurements configurations.

5.3.1 Topology Estimation in Substations - External Measurements

The work on a breaker status reconstruction performed in section 5.1, will be verified here on the split bus problem. For the proof of concept, a representative substation sample was chosen. It is the bus 15 from the IEEE RTS 24-bus system, which is now considered to be a substation with a breaker arrangement as shown in Figure 5.12. The substation scheme is double bus/double breaker plus a breaker-and-a-half-breaker for the connection of the generator.

If the status of each particular breaker is observed, a large number of possible substation configurations may be considered ($2^7 = 128$). However, one assumes that power flow meters are located outside of the substation, i.e. only external power measurements are available. As discussed earlier, it is not possible to identify all the breaker statuses only with external measurements, but a group of different status configurations which result in identical external measurements. Therefore, a single representative from a group of indistinguishable topologies is considered, so only 35 topologies are taken into account for the substation 15. In general, such an external split bus problem is defined considering the following assumptions:

- ▷ A substation is assumed to be a black box,
- ▷ Only external information is used; i.e. measurements of power flows through circuit breakers are not available,

- ▷ Multiple topologies which are indistinguishable from external measurements are merged to a single representative topology,
- ▷ The objective is to determine the substation topology which implies a global network topology consistent with the external measurements.

For the purpose of training and testing the autoencoders, the database was generated by superposing a noise signal over the load flow results. Settings of the network parameters (load curve and generation rates) are identical to the ones described in section 5.1.

5.3.1.1 Global vs. Local Self-tuning Model for the Split Bus Problem

In order to verify the conclusions derived from a single breaker status identification, the self-tuning auto-associative methodology will now be tested on the split bus problem. For the purpose of breaker status reconstruction, the self-tuning model searched between the combinations of open or closed breakers statuses. Now, in the case of substation 15, it will search between the 35 possible topologies.

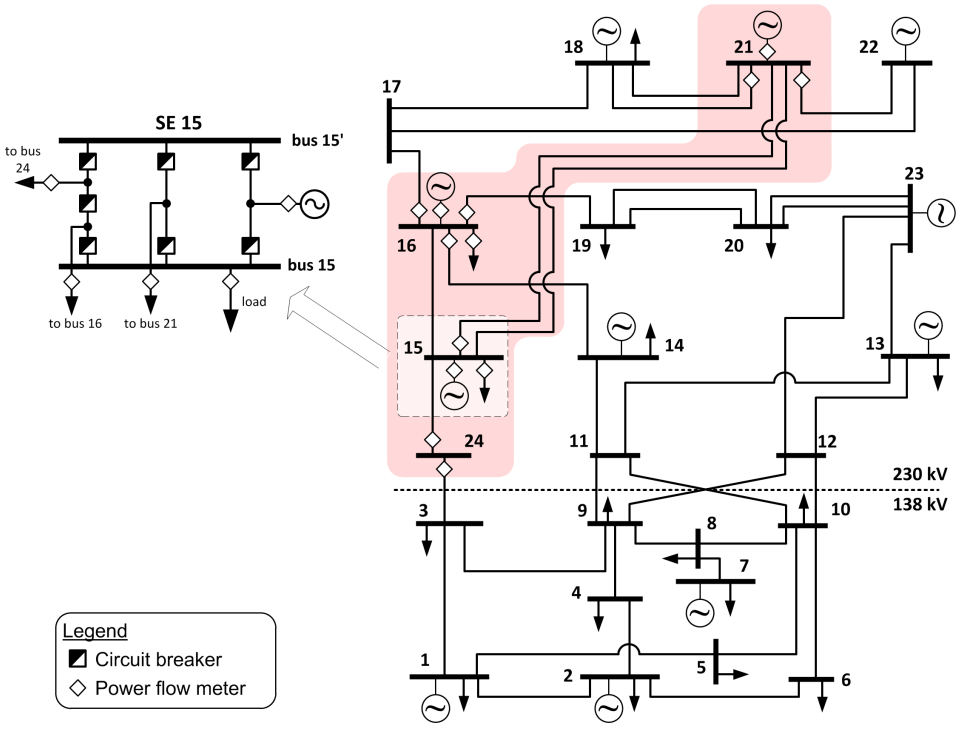


Figure 5.12: Connection scheme of substation 15. Red-shade denotes the area of measurement consideration for local autoencoder.

The global concept takes into account all the available power measurements, resulting in large autoencoder size of 125-60-125. Once again, the local model should test the hypothesis that a change in a substation topology has a strong local character. In that regard, a small local

autoencoder (33-20-33), collecting only local information (Figure 5.12), is expected to perform, at least as efficiently as an autoencoder trained with measurement data of the entire system.

Results of the entire substation topology reconstruction for global and local autoencoder are presented in Table 5.12. Both autoencoders take into account only active and reactive power measurements. It is very significant that a local autoencoder, which is almost 4 times smaller, is able to perform almost as efficiently as the global one (79.25% against 84.89%). Although this result seems logical, the local model is expected to be even superior due to a more focused and less demanding training process. The reason why the local model still lags behind may be due to a choice of most relevant inputs, which are determined by guess, i.e. by an engineering common sense. Obviously, some information about the topology variable of the substation 15 remains outside of the red-shaded area in Figure 5.12. In that regard, a more sophisticated method should be proposed in order to extract the maximum information from the available measurement set, i.e. to select input measurements which are the most informative about the substation topology. This issue addressed in detail in Chapter 6.

Table 5.12: Reconstruction efficiency of substation 15 topology with a global and local self-tuning autoencoder.

	Wrong	**Correct**	**Efficiency**
Global	1511	8489	84.89%
Local	2075	7925	79.25%

5.3.1.2 Competitive Model and Topology Cells for the Split Bus Problem

Section 5.1 has already discussed some deficiencies of the self-tuning model and proposed an alternative method called competitive, which has been shown to be substantially superior due to multiple aspects. For example, in order to identify a topology in the real time the self-tuning model requires an optimization method to reconstruct the unknown values. In the case of abundant topology variants, a significant challenge emerges for such an optimization in real time environment. Accordingly, it may become a relevant trade-off between the identification accuracy and the computational requirements.

On the other hand, the competitive scheme provides topology estimation practically instantaneously and is also expected to be more efficient. Therefore, the intention is to adopt the same concept in the split bus problem. This is done by approaching the topology estimation problem from the pattern recognition point of view. Hence, the following steps are proposed to be taken:

1. All the considerable topologies will be identified from the available measurement dataset.

2. For each topology, a single autoencoder will be trained with the corresponding dataset.

3. In real time, the relevant measurement vector will be presented to all autoencoders. The one which appears coherent with the data, i.e. responds with the smallest input-output error, "wins the competition" and indicates the estimated substation topology.

As is illustrated in Figure 5.4, each topology defines a different manifold in the space of power measurements, which is captured through the training concept from each particular autoencoder. When a new measurements vector is presented to all autoencoders, the one which corresponds to the closest manifold is expected to respond with the smallest input-output vector, and indicate the most probable topology along the way.

For a single breaker, a competitive scheme is reduced to the simplest level, i.e. only two options are possible: open or closed breaker. For a group of breakers in a substation, a competitive scheme may be much more complex due to numerous topology concepts. For the purpose of concept generalization, the term ***Topology Cell*** (**TC**) is defined which refers to a topology entity defined by a group of breakers considered by the same competitive scheme. Hence, the most primitive topology cell can be a single breaker with 2 corresponding autoencoders. The other end-case would be when a topology cell representing an entire system which can take numerous topologies.

Moreover, a topology cell of the observed substation 15 will correspond to 35 competitive autoencoders for 35 topologies distinguishable from external measurements. Results in Table 5.13 indicate an enormous improvement in comparison to the self-tuning autoencoders. Only in 14 test case scenarios out of 10000, competitive autoencoders failed in identification of the entire substation topology. Moreover, it is impressive that only 26 incident active and reactive power measurements, external to the substation, are enough for such an efficient topology estimation. Therefore, in comparison with the local self-tuning model, the efficiency in topology estimation of the substation 15 is improved from 79.25% to an impressive 99.86% (Table 5.12).

Table 5.13: Results of competitive substation topology identification for substations 1, 2, 9, 10, 13, 15 and 16, considering only external measurements. Topology cell is associated to an entire substation and the number of required autoencoders is equal to the number of topologies, distinctive from external measurements.

	Model parameters			**Substation topology reconstruction**		
Substation	Nº TC	Nº AANN	Nº topologies	Wrong	Correct	**Efficiency**
SE 1	1	21	21	0	10000	**100.00%**
SE 2	1	21	21	0	10000	**100.00%**
SE 9	1	52	52	24	9976	**99.76%**
SE 10	1	52	52	3	9997	**99.97%**
SE 13	1	21	21	0	10000	**100.00%**
SE 15	1	21	21	14	9986	**99.86%**
SE 16	1	37	37	0	10000	**100.00%**

The proposed competitive scheme is now tested for some other representative substations in the 24-bus system. Due to limited number of topologies which can be distinguished from external measurements, only substations with more complex configurations are considered, i.e. with 3 or more incident lines. For all the substations, the internal circuit breaker arrangement is assumed to belong to a double-bus/double-breaker scheme, with the addition of a bus tie breaker, as depicted in Figure 5.10. A single topology cell is defined for each substation, so the number of autoencoders is equal to the number of topologies that a substation can take, regarding the complexity of the substation arrangement, i.e. number of incident lines, loads and generators.

Results presented in Table 5.13 show an impressively high topology estimation efficiency for all the substations observed. Topology reconstruction is impeccable in most of the cases, while a few failures are noted in more complex substations. In addition, computational requirements for this topology estimation procedure are practically negligible. These results are very significant since they prove the applicability of the novel competitive auto-associative topology estimator, which is able to be very efficient, even without any prior knowledge about breaker statuses and without any measurements within the substations.

5.3.2 Topology Estimation in Substations - Internal Measurements

This section intends to test the competitive auto-associative methodology in scenarios when measurements internal to a substation are also available. More specifically, the previous section assumed that power flow meters are located only outside of the substation. Here, measurement configuration is defined in a way that each circuit breaker has associated a meter that measures power flow through it (right side of Figure 5.10).

Table 5.14: Results of competitive topology identification for substations 1, 2, 9, 10, 13, 15 and 16 considering the internal power flow measurements. It is notable that the efficiency method significantly depends on substation complexity.

	Model parameters			**Substation topology reconstruction**		
Substation	Nº TC	Nº AANN	Nº topologies	Wrong	Correct	**Efficiency**
SE 1	1	136	136	193	9807	**98.07%**
SE 2	1	136	136	277	9723	**97.23%**
SE 9	1	602	602	4020	5980	**59.80%**
SE 10	1	602	602	3109	6891	**68.91%**
SE 13	1	136	136	1889	8111	**81.11%**
SE 15	1	136	136	468	9532	**95.32%**
SE 16	1	288	288	704	9296	**92.96%**

The method is tested on the same substation cases as in the previous section, but due to availability of internal power flows, a number of topology variants was drastically increased. In

addition, there is no prior knowledge about the circuit breaker statuses. Results are presented in Table 5.14. Since a single topology cell is defined over the entire substation, the number of required autoencoders is equal to the number of expected topologies. For more complex substation arrangements, the abundance of autoencoders may become concerning due to feasibility issues in more realistic circumstances. However, this should not be a big issue, since local autoencoders are quite reduced in size, so the training process is not demanding at all and it is also performed offline. The reconstruction/estimation process which is run online is always straightforward, independent of the number of autoencoders. On the other hand, one should be much more concerned with a drop in efficiency which is substantial with an increase in substation complexity. This is particularly notable in the results for substations 9 and 10, which, in comparison to other substations, require significantly more autoencoders.

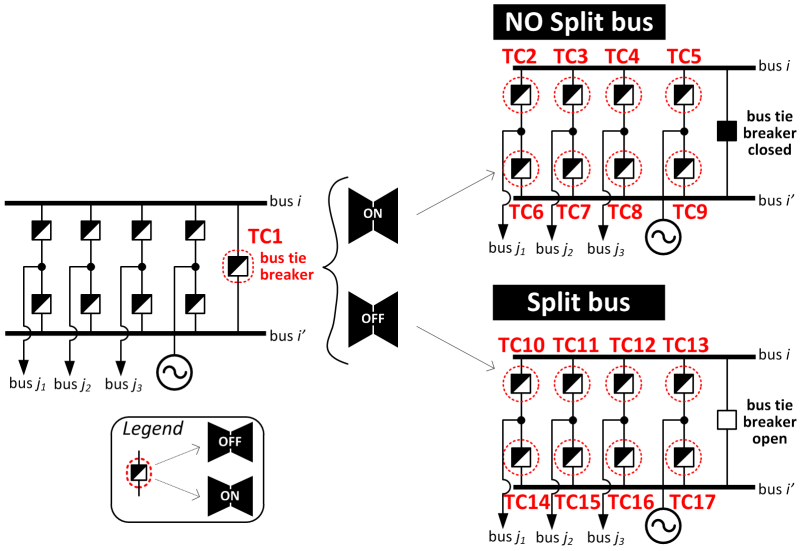


Figure 5.13: Decentralized scheme of competitive autoencoders for substation topology reconstruction using the example of a sample substation with double-bus/double-breaker configuration. For 17 topology cells {TC1,...,TC17} there are 34 autoencoders in charge of identification of their topologies.

In the performed test, a single topology cell is associated to an entire substation. In regard to this, it can be observed that the arrangement of topology cells directly defines the number of required autoencoders, and accordingly, is also expected to impact the efficiency of the method. Therefore, the estimation performances are decided to be tested for the other end-case of topology cells arrangement, i.e. a single topology cell is associated to each circuit breaker, instead of the entire substation. Also, one topology cell is set to be in charge of the bus-tie breaker, so the cases when the bus is split can be distinguished from the cases when only a single-bus scheme is considered. This dispersed TC arrangement is illustrated for a sample substation with a double-bus/double-breaker configuration in Figure 5.13.

The results of a substation topology estimation, with an alternative TC arrangement (without changing any other method parameters or input data) are presented in Table 5.15. When compared

to results from Table 5.14, it is important to note the enormous saving in the number of required autoencoders. For example, in the cases of most complex substations 9 and 10, the number of autoencoders is reduced approximately 12 times (from 602 to 50). The efficiency of topology estimation is improved for some substations, and a drop in efficiency is not significant for the others. A conclusion can be derived that a dispersion of TC through a substation results in drastic saving in computational and memory requirements of the training process, while the estimation efficiency is on average kept on the same level.

Table 5.15: Results of topology estimation, considering internal measurements, with the dispersed competitive autoencoders scheme. In comparison to results of a single topology cell from Table 5.14, the efficiency is improved in some cases but with substantially fewer number of autoencoders.

	Model parameters			**Substation topology reconstruction**		
Substation	№ TC	№ AANNs	№ topologies	Wrong	Correct	**Efficiency**
SE 1	21	42	136	899	9101	**91.01%**
SE 2	21	42	136	847	9153	**91.53%**
SE 9	25	50	602	3669	6331	**63.31%**
SE 10	25	50	602	2849	7151	**71.51%**
SE 13	21	42	136	2921	7079	**70.79%**
SE 15	21	42	136	1667	8333	**83.31%**
SE 16	25	50	288	2556	7444	**74.44%**

In summary, the tests performed in this section revealed the importance of topology cells arrangement in performances of the proposed topology estimation method. In other words, a point is raised that the competitive autoencoder scheme needs to be optimally defined according to dependencies among topology variables. A principal idea is to divide the topology estimation problem into multiple problems of smaller size, with an intention to minimize the number of required competitive autoencoders and to also enhance the estimation reliability. Therefore, it can be concluded that a proper method for TC arrangement optimization would be of key interest for the upgrade towards an industrially-acceptable methodology. A concept that provides the answer to this challenge is proposed in the subsequent chapter.

5.4 Conclusions Derived - Auto-Associative Topology Estimator

This section proposes a novel power system topology estimator based on a mosaic of local Auto-Associative Neural Networks (AANN or Autoencoders). The main idea is to use autoencoders to capture the mutual information between the electric and the binary topological

variables associated to the circuit breaker statuses. The main issues on which the proposed methodology is established on can be summarized as the following:

- Information on breaker statuses in a power network is embedded in values of electric analog variables of the system.
- An autoencoder is able to extract information about the system topology and capture the manifold of available analog electric measurements through the training process.
- The information extracted from the dataset is stored in the autoencoder weights during the training process, which is performed offline.
- The trained autoencoder is used in real time for the reconstruction (estimation) of unknown breaker status, related to available measurement vector and the information deposited in its weights.

In comparison to the existing topology estimation methodologies, the proposed concept can be characterized as unconventional, *a priori* method that is independent to any classical estimator. First of all, the method resides on the hypothesis that the information on a breaker status is hidden in the values of electric analog variables of the system. Also, another hypothesis of key importance is that a particular network topology will constrain the solutions of Kirchhoff Laws to a specific pattern. This being so, one approaches the topology estimation problem from the pattern recognition point of view. Therefore, if both analog measurements and digital breaker status values are submitted to the autoencoder, it is expected that it may learn the affiliation of each binary code with a corresponding pattern defined by a set of analog measurements. This pattern is characteristic for each topology state and corresponds to the manifold of possible power flow solutions conditioned by each topology.

This chapter developed a comparison between a global centralized approach and the adoption of a mosaic of local estimators in a decentralized fashion, very parallel to the concept developed in Chapter 4.

The first approach discussed has used all the available analog (power and voltage measurements) and relevant binary electric signals (breaker status) for the training of a single autoencoder. The reconstruction of some of the unknown (or bad) signals is performed in real time by a constrained search optimization (using EPSO). This model can be characterized as global since all the available data about the system, available at the SCADA, are being considered at once. The preliminary tests showed very promising results. However, the remaining concern is the applicability of such method to larger problems. Although training is to be performed offline, the exponential growth in the number of weights needed to be trained threatens the feasibility of an accurate neural network tuning. Furthermore, the computational costs of the constrained search (optimization procedure) may become insuperable for more complex topology arrangements. Finally, it is generally more difficult for the global model to keep the required level of reliability in diverse scenarios, i.e. it can barely respond to challenges like: optimal response to all the possible

operational scenarios, robustness to gross errors or flexibility to structural changes of a power system.

Generally, when observing the effects of a breaker status change on any type of electric variables, it can be noted that they are limited to a very local area. This raises the hypothesis that the reconstruction of a breaker status signal could be performed considering only some local power and voltage measurements instead of all data available about the system. Therefore, a general model of a mosaic of local autoencoders has been established: local autoencoders are distributed throughout the system in a way that each one is in charge of a particular topology cell - a single breaker or a group of breakers. This idea of disseminating the localized topology estimators makes a substantial step towards the decentralization of the estimation processes. This issue is of key significance for building an environment suitable for the coming modern power system concepts. In general, the benefits of decentralization and localization of the particular model are multiple:

- ▷ drastic reduction of the training requirements,
- ▷ robustness to gross errors,
- ▷ enhancement of estimation efficiency,
- ▷ flexibility regarding operational and structural changes in an electric power system.

The model of the constrained search over the autoencoder inputs is identified in this thesis as the self-tuning. It has already been used in Chapter 4 for the reconstruction and correction of analog measurements. With an appropriate adjustment of the objective function (Correntropy) it is used in this chapter very efficiently for the reconstruction of binary breaker status signals. The computational costs demanded by the constrained search are considered to be the main drawback of this method.

Motivated by this observation, an alternative approach to this problem was investigated. Each network topology constrains the data to a specific pattern, so the estimation of topology may be perceived as a pattern recognition problem. This has resulted in a scheme denoted as competitive autoencoders, adopted from [150] and applied for the first time in this thesis for the estimation of power system topology. For each topology possible in a given topology cell, an autoencoder is trained with the relevant dataset. After capturing the manifolds described by each topology, a set of autoencoders is organized in a real time competitive scheme: the measurement vector acquired by the SCADA is submitted to all the autoencoders and the one which displays the most coherent behavior (responds with the lowest input-output error) indicates that topology associated to it is the most probable one. In all the performed tests, the competitive scheme significantly outperformed the self-tuning model, regarding both topology estimation efficiency and computational requirements.

The novel scheme of local competitive autoencoders was also tested in complex breaker arrangements such as substations. Two types of measurement configurations were considered in particular: measurements external to substations and internal measurements of the power flow associated to each breaker. The method is shown to be high efficient in both measurement

configurations. An extreme value of the method is especially recognized in specific cases when available measurement sets are scarce and do not relate to the interior scheme of substations. These circumstances are expected to occur more frequently in distribution grids, where nowadays the need for topology estimation becomes more and more demanding.

In summary, the general contributions brought by the proposed power system topology estimator are multiple:

- ◇ The topology estimation problem is perceived from information theoretic perspective: the information about breaker statuses in a power network is embedded in the pattern of electric variables.
- ◇ The establishment of a general decentralized scheme of local autoencoders: this architecture brings about multiple advantageous properties regarding suitability with modern power system concepts.
- ◇ The competitive topology estimation scheme is extremely suitable for the real-time environment - it estimates the topology practically instantaneously, with a negligible computational cost.
- ◇ The contributions of the proposed topology estimator are particularly valuable for the most demanding cases, when the metering infrastructure is scarce and does not provide enough redundancy even for the observability of the system.

The proposed topology estimator is built to be compliant with the present trends in the evolution of the distributed state estimation ([57], [58], [61]), in a direction compatible with the modern power system concepts, namely smart grids and micro-grids.

All the models were built and tested assuming a given architecture for the data collection. The definition of input data sets was based on engineering judgment. A question remains unanswered: is it resorted to the best possible set, in terms of information content, when the topology is tried to be derived out of a collection of measurements? The following chapter provides an answer to this challenge.

Chapter 6

Statistical Observability in Power Systems

In the process of power system state estimation, apart from topology processing, bad data processing and state estimation (in the narrow sense), there is one more issue to be addressed: the observability analysis. Put simply, the network observability analysis will determine, according to a given set of measurements and their locations, if a unique solution can be found for the system state. In other words, a network with n state variables and m observed measurements, where $m > n$, is declared to be non-observable if a system with n independent equations cannot be built. Alternatively, if a system of equations, including all the state variables, can be solved, then the network is considered to be observable. Hence, a classical definition of observability is conceptualized as binary (or Boolean): the system either is or is not observable.

Subsequently, the estimation process for non-observable networks can only be solved for an observable subsystem - the so-called observable island, which corresponds to $n' < n$ state variables. Due to insufficiency of the available observations, one refers to non-observable state variables $(n - n')$ and the associated structural elements of the network. In that case, the traditional approach is primarily concerned with observability restoration. The mostly used technique for this purpose is generation of pseudo-measurements.

In Chapters 4 and 5, a novel way of processing of electric variables has been proposed. The proposed method performs measurement screening in order to identify bad data and reconstruct their assumed true values. The main reconstruction tool is the auto-associative neural network that extracts information from available variables during training process and stores it in its weights. The efficiency of this extraction process depends on many method parameters, but one of most important factors will be the selection of input data vector. In Chapters 4 and 5, the selection of a set of local measurements was based on a simple engineering judgment: the measurements which are electrically closest to the variables of interest are chosen as the most relevant. This criterion seems logical and is probably not so far from the optimum. However, it evidently lacks the theoretical justification which could provide the appropriate upgrade and complete the proposed approach. The main question that addresses this problem is - *Which input set of electric variables is the most*

informative about a particular topology? This chapter is a first attempt to provide a partial answer to this question and results with a new way of reasoning about power system observability.

The new perspective given in this chapter uses the information theoretic descriptor called Mutual Information, in order to detect the most informative measurements for estimation of some variables of interest. MI generally measures a decrease in uncertainty of a variable by observing another variable. The proposed concept offers a novel perspective in observability of an electric variable. By measuring the information that one electric variable contains about another, it also provides a quantification of the observability degree. In other words, the intention is to measure the relevance of measurements for observing other electric variables.

Unlike the classical concept defined on $\{0,1\}$, this representation by mutual information expresses observability within the range $[0,1]$. Hence, instead of asking if a variable is observable or not, one will be concerned about how much of it is observable. In fact, the gain in information about a variable by knowing another variable will be measured. This way, the new concept intends to deal with observability in a way free from the algebraic requirements for solving a system of equations – therefore, it is called *Statistical Observability*. It must be underlined that this concept is not proposed to substitute the traditional model: it will serve as a complement and support, since it enables an accurate reconstruction of missing values or generation of pseudo-measurements which contribute to system observability. Moreover, the concept of observability, explored in this chapter, does not refer to *system observability* but to *variable observability*. It clearly is a way of generalizing the otherwise well established concept of observability, having in mind that a system can only be observed if all state variables can be known, i.e. observed. Although understanding the possible objections from researchers that would like to reserve the term *observability* as a system characteristic, this thesis nevertheless will adopt the definition of *statistical observability of variables*. In the text, the expression *statistical observability* will be used with this meaning.

The work presented within this chapter is particularly concerned with topological observability. Mutual information is being used as a criterion for definition of the mosaic of local autoencoders, which are expected to outperform the manually-defined autoencoders from Chapter 5. The same concept is also assessed as suitable to determine an optimal level of decentralization, i.e. size of the topology cells, as proposed in Chapter 5. In any respect, the statistical observability concept is straightforward for any other type of variables and generally contributes to automation and efficiency of the established models.

6.1 Mutual Information - a General Overview

Mutual information measures a decrease in uncertainty of a random variable by observing another variable. In other words, it measures the amount of information the two variables contain about each other (this may be generalized to multiple variables or vectors).

In general, if two random variables X and Y are independent, it means that each variable does not contain any information about the other one, so joint entropy $H(X)$ is equal to the sum of particular entropies (equation (3.11)). Alternatively, if any dependency exists between them,

joint entropy is decreased by ***Mutual Information*** (**MI**) $I(X,Y)$ between X and Y, or information shared by X and Y. Hence, MI quantifies the intersection of $H(X)$ and $H(Y)$, or in other words, it quantifies the reduction of uncertainty in X, after observing Y: $I(X,Y) = H(X) - H(X|Y)$. Also, the same result can be obtained from the other side: $I(X,Y) = H(Y) - H(Y|X)$. It is important to say that MI is always greater or equal to zero and is upper bounded by minimum of the two entropies $\min\{H(X), H(Y)\}$.

Mutual information can also be defined as a special case of ***Kullback–Leibler divergence***. Let $p(x)$ and $q(x)$ be two different probability densities. Kullback-Leibler divergence between them is defined, for the continuous and the discrete case, respectively:

$$D_{KL}(p\|q) = \int p(x) \log\left(\frac{p(x)}{q(x)}\right) dx \tag{6.1}$$

$$D_{KL}(p\|q) = \sum_{x} p(x) \log\left(\frac{p(x)}{q(x)}\right) \tag{6.2}$$

Therefore, Shannon's Mutual information is obtained as the Kullback–Leibler divergence between the joint PDF $p(X,Y)$ and product of the marginals $p(X)p(Y)$ [115]:

$$I(X,Y) = D_{KL}(p(X,Y)\|p(X)p(Y)) = \iint p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right) dxdy \tag{6.3}$$

For the two discrete variables X and Y it holds:

$$I(X,Y) = \sum_{j=1}^{N}\sum_{i=1}^{N} p(x_i,y_j) \log\frac{p(x_i \mid y_j)}{p(x_i)} = \sum_{j=1}^{N}\sum_{i=1}^{N} p(x_i,y_j) \log\frac{p(x_i,y_j)}{p(x_i)p(y_j)} \tag{6.4}$$

It is significant to observe that if the product and the joint are equal, then variables are independent, thus the Mutual Information is zero and the equation (3.11) holds.

Now, if Renyi's entropy is considered, the corresponding Renyi's divergence or order-α divergence between $p(x)$ and $q(x)$ is defined as:

$$D_{\alpha}(p\|q) = \frac{1}{\alpha - 1} \log \int p(x) \left(\frac{p(x)}{q(x)}\right)^{\alpha-1} dx \tag{6.5}$$

The Renyi's order-α Mutual Information can be derived as divergence between the joint distribution and the product of marginal distributions $p(X)p(Y)$:

$$I_{\alpha}(X,Y) = D_{\alpha}\left(p(X,Y)\|p(X)p(Y)\right) = \frac{1}{\alpha - 1} \log \iint p(x,y) \left(\frac{p(x,y)}{p(x)p(y)}\right)^{\alpha-1} dxdy \tag{6.6}$$

For the Renyi's quadratic mutual information between X and Y, it holds:

$$I_2(X,Y) = \log \iint \frac{p^2(x,y)}{p(x)p(y)} dxdy \tag{6.7}$$

In the case of two discrete variables X and Y, Renyi's quadratic mutual information is calculated

as:

$$I_2(X,Y) = \log \sum_j \sum_i \frac{p^2(x_i, y_j)}{p(x_i)p(y_j)} \tag{6.8}$$

Mutual information can also be calculated for the multivariate case. Let X be the n-dimensional random vector X, with joint distribution $p_X()$ and p_{X_j} as marginal density of the j-th component of X. Then quadratic Renyi's mutual information between the components is evaluated as [115]:

$$I_2(X) = \log \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{p_X^2(x_1, \ldots, x_n)}{\prod_{j=1}^{n} p_{x_j}(x_j)} dx_1 \ldots dx_n \tag{6.9}$$

Among the presented definitions of Mutual Information, the quadratic Renyi's mutual information will be adopted for the purpose of consistency with other concepts used in this thesis. In any case, there are not specific benefits from using Renyi's definition, so Shannon's definition could also be used without a significant discrepancy between the results.

6.2 Mutual Information for Local Autoencoder Determination

Chapter 5 of this thesis proposed a power system topology estimator, based on competitive auto-associative neural networks with a decentralized local architecture. The main hypothesis this estimator is built on is: the network topology conditions the shape of the manifold which holds a continuous set of solutions of the equations and defines the state variables. Therefore, the electric variables, like the measurements of power or voltage, are expected to contain the information about the underlying network topology. One could say that the topology information is hidden in the values of the available measurements. This hidden information is proposed to be extracted in Chapter 5 by the auto-associative neural networks.

When addressing this problem, an important question still remained unanswered: *Which input measurements are most informative about the particular topology?* The heuristic answer to this question, offered in Chapter 5, cannot be satisfactory: the initial criterion for definition of mosaic of local autoencoders was simply based on topological adjacency, i.e. on the belief that the impact of changing a breaker status is retained mostly in incident measurements.

In this section, one intends to offer a more theoretically justifiable answer to this question via systematic identification of variables which are most affected by a change in the breaker status. In other words, the objective is to evaluate the informational content the available measurements provide about a particular topology cell and accordingly define the autoencoders which are expected to perform more efficiently. A theoretical background suitable for this purpose is found again in the Information Theory, i.e. in the Mutual Information criterion, referred to in the previous section.

Among many feature selection methods available in the literature, the reasons for selecting MI are multiple. Most of the existing methods are based on sensitivity analysis, linear transformations or principal component analysis. Unlike most of the well-established methods, MI possesses two key properties:

- ▷ Ability to measure an arbitrary relationship between variables,
- ▷ Invariance under an invertible, linear space transformations.

The first property refers to the fact that information-theoretic methods also consider nonlinear relationships between random variables, unlike the conventional descriptors that use second-order statistics for comparison of linear relationship between random variables.

By recognizing these properties, one intends to use mutual information between topology variables and measurements in order to build a more efficient auto-associative topology estimator. In fact, the selection of measurements for an auto-associative topology estimator is nothing more than a selection of features for a classifier. Accordingly, it must be noted that the adoption of MI for feature selection is not novel. The first proposal of this type, published in [159], uses MI as a criterion for selecting features for a general supervised neural network learning. A more detailed overview about the adoption of MI, as a feature selection criterion, is available in Appendix B.2.

The intention of feature selection, in general, is to evaluate a set of candidate features and select a subset that is most informative for a particular classification problem. The benefit of such pre-processing is the satisfaction of trade-offs between training requirements and classification efficiency.

In general, the feature selection problem can be formulated as the following: if an initial set F of n features is given, the subset $S \subset F$ with $k < n$ features that is maximally informative about the classes C is intended to be found. In the framework of Information Theory, the term *maximally informative* is incorporated through ***Maximum Mutual Information*** (**MMI**) criterion between features and the class - $I(C,S)$. An equivalent criterion can be determined with minimum conditional entropy of the class by knowing the subset S of features - $H(C|S)$.

Therefore, aim of the MI feature selection is to find the subset S of features which results with maximum value of MI with the considerable classes. It is very important to understand that theoretically, one should take into account the interaction between groups of attributes and not only maximize the criterion in respect to a single feature. In other words, the group of k best features does not necessary make the best group with k features. In addition, a group of several features acting simultaneously can appear to be very relevant while the individual features alone do not contribute at all. Therefore, an ideal feature selection algorithm should take into account multivariate mutual information $I(C,S)$ for all subsets $S \subset F$. Unfortunately, the criterion is barely implementable on that way due to the following computational burdens:

- ▷ The computing effort needed for multivariate mutual information processing exponentially grows with an increase in the number of features (see the equation (6.9)),
- ▷ The consideration of all possible subsets requires $\binom{n}{k}$ calculations of MI, where n is total number of features and k ($k < n$) is a predefined number of features in a subset S.

Therefore, certain criterion approximations need to be done, in order to satisfy the trade-off between computational requirements and relevancy of the criterion. For that purpose, multiple solutions were proposed in the relevant literature, as Appendix B.2 explains in detail. All the

proposed criteria intend keep relevance of the criterion despite introduced approximations. This is principally obtained by maximization of information provided by individual features, and meanwhile minimization of redundant information between the features.

In the particular auto-associative state estimation, from Chapter 5, one needs to be aware that efficiency of the method depends also on the autoencoder architecture and efficiency of the training process, so it is not only about quality of the input data. According to the performed preliminary case studies, it can be derived that quantification of the redundant information among measurements spends precious computational efforts, without any significant gain in estimation efficiency. Therefore, it can be concluded that consideration of individual features is only enough for building an efficient mosaic of autoencoders. In other words, the proposed MI formulation for measurements selection ignores redundant information that multiple measurements contain about a topology variable and considers only individual contributions. Hence it is only necessary to calculate MI for all the measurements, and select the k best according to the MI ranking.

The Renyi's quadratic Mutual Information from (6.8) is adopted for calculation of MI between topology variable, associated to the topology cell T, and the particular measurement z. If one assumes them to be discrete variables and estimates joint distribution $p(z,t)$ and marginals $p(t)$ and $p(z)$, information contained in the measurement z about the topology variable T can be calculated as:

$$I(T,z) = \log \sum_j \sum_i \frac{p(z_i,t_j)^2}{p(z_i)p(t_j)} \tag{6.10}$$

The following sections report some case studies and analyze abilities and contributions of using Maximum Mutual Information (MMI), as a criterion for defining competitive local autoencoders for the topology estimation problem.

6.2.1 Case Study - a Single Breaker Status

In order to demonstrate the benefits of adopting the new measurement selection criterion, the method will be tested on the same case study analyzed in Chapter 5 with 10 breakers added in the 24-bus system (Figure 5.1). The results are compared with the ones obtained by autoencoders determined according to engineering judgment criterion. There is a reason to expect that measurements which contain high mutual information are electrically close to the breakers. However, the mutual information is expected to be a more relevant criterion and thus contribute to enhancement of efficiency in topology estimation.

The mutual information index between the active and reactive power measurements and the topology variables (status of each switch) was calculated for the same database used in section 5.2. Figures 6.1 and 6.2 illustrate a map of mutual information for active and reactive power measurements for the most representative breakers 6 and 9. The edge colors are associated with power flow measurements and node color with power injection. One can say that these figures represent an ***Observability Map*** of a breaker status, in the information theoretic sense.

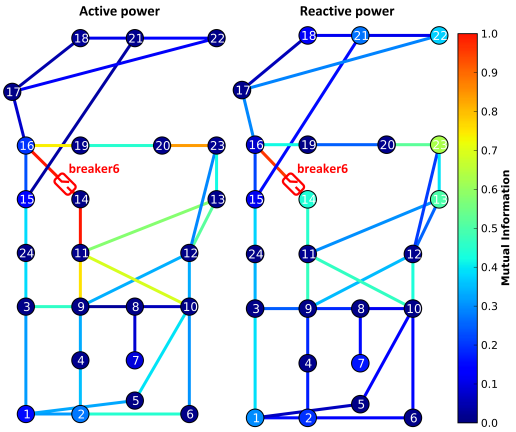


Figure 6.1: Mutual information maps for active (left) and reactive power (right) regarding breaker 6. Edge color corresponds to power flow measurements, while node color is associated with power injection. There is a significant spread of mutual information with the breaker 6, unlike the breaker 9 (Figure 6.2). Maximum MI is contained in the active and reactive power flow through line 14-16.

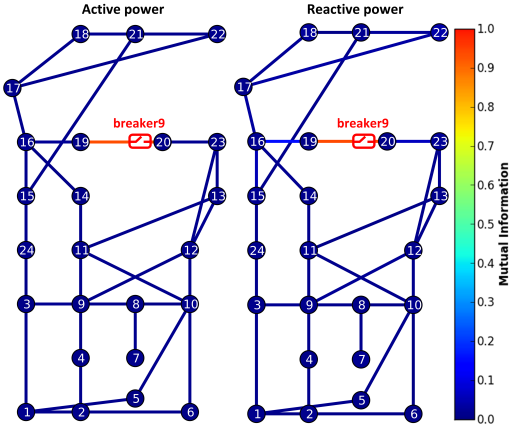


Figure 6.2: Mutual information maps for active (left) and reactive power (right) regarding breaker 9. Edge color corresponds to power flow measurements, while node color is associated with power injection. Information about the breaker is limited almost exclusively to the active/reactive power flow in line 19-20.

It is significant to note the high spread of information about breaker 6, while in the case of breaker 9, the information is limited to quite a local area, i.e. it is almost exclusively kept in the power flow through the breaker (line 19-20). In addition, a high similarity is evident between MI maps of active and reactive power. However, it is particularly interesting to note the high informational content about breaker 6 in the reactive power injection in node 14. This result is justifiable if one knows that a synchronous condenser is connected to bus 14 and is set for the purpose of voltage regulation.

It is also logical that measurements of power flow through the breaker contain the largest

mutual information with the breaker status. Answer to the question why the active/reactive power flow measurement through the breaker line has the most mutual information with the particular breaker status lays in the following relation:

- $P_{i-j}^{flow} = 0 \quad \Rightarrow \quad$ breaker $i-j$ is open;
- $P_{i-j}^{flow} > 0 \quad \Rightarrow \quad$ breaker $i-j$ is closed.

where P_{i-j}^{flow} and Q_{i-j}^{flow} are the active and reactive power flows in the line $i-j$.

Table 6.1: Mutual information between breaker statuses and measurements of active and reactive power flows in particular breaker lines. It is important to note that for most of the breakers, at least active or reactive power flow through the breaker contains maximum information about the breaker status, i.e. contains maximum possible value of mutual information.

Breaker (T)	measurement (z)	$MI(z,T)$
1	P_{1-2}^{flow}	0.963
	Q_{1-2}^{flow}	1.000
2	P_{3-9}^{flow}	0.989
	Q_{3-9}^{flow}	1.000
3	P_{5-10}^{flow}	0.954
	Q_{5-10}^{flow}	0.984
4	P_{11-13}^{flow}	1.000
	Q_{11-13}^{flow}	0.887
5	P_{12-13}^{flow}	1.000
	Q_{12-13}^{flow}	0.916
6	P_{14-16}^{flow}	1.000
	Q_{14-16}^{flow}	0.977
7	P_{15-21}^{flow}	1.000
	Q_{15-21}^{flow}	0.952
8	P_{17-18}^{flow}	0.995
	Q_{17-18}^{flow}	0.971
9	P_{19-20}^{flow}	0.971
	Q_{19-20}^{flow}	0.943
10	P_{21-22}^{flow}	1.000
	Q_{21-22}^{flow}	0.986

Therefore, these measurements mostly contain a direct information about the breaker status. This is also evident from the value of mutual information that is equal to 1.0 for most of the power flows through the breaker. More specifically, the variable that defines the breaker status can take values of 0 or 1 with the same probability and the entropy of such an event is equal to 1. Since mutual information is upper bounded by the minimum entropy of the two variables, theoretically maximum MI value in this case is 1. Therefore, if a measurement contain MI of 1, it means it contains a direct information about the breaker status. If knowing such variable, there is no need for some sophisticated breaker status reconstruction method, but only a heuristic method is sufficient, as demonstrated in section 5.2.1. However, as it is discussed before, it is possible that power flow is zero although the breaker is closed. These scenarios are expected to deteriorate the informational content about the breaker status to a value below 1.

For all the 10 breakers, Table 6.1 shows values of mutual information between each particular breaker and associated measurements of active and reactive power flow through them. It can be noted that, for each breaker, at least one of the active or reactive power flow measurements reaches the maximum value of 1. Exceptions are the breakers 3, 8 and 9, for which ambiguous power flow scenarios are probable to occur.

Table 6.2: Comparison of a single breaker status reconstruction results, obtained by two different input measurement sets. One is defined by the engineering judgment and the other by the MI criterion. It can be observed that the MI criterion is able to contribute to breaker status reconstruction efficiency.

Breaker	**Engineering judgment**			**Mutual Informations**		
	Wrong	Correct	Efficiency	Wrong	Correct	Efficiency
1	0	10000	100.00%	0	10000	100.00%
2	0	10000	100.00%	0	10000	100.00%
3	17	9983	99.83%	2	9998	99.98%
4	0	10000	100.00%	0	10000	100.00%
5	0	10000	100.00%	0	10000	100.00%
6	0	10000	100.00%	0	10000	100.00%
7	0	10000	100.00%	0	10000	100.00%
8	7	9993	99.93%	5	9995	99.95%
9	0	10000	100.00%	0	10000	100.00%
10	0	10000	100.00%	0	10000	100.00%
Total	24	99976	**99.98%**	7	99993	**99.99%**

According to the presented MI maps, one can already observe that there is no one-to-one correspondence between the geometrical adjacency with the breaker and mutual information. Thus, the results produced by MI and engineering judgment are expected to be, at least, somewhat

different. In order to justify the hypothesis, 10 couples of competitive local autoencoders are built according to mutual information criterion and used for the estimation of 10 breakers statuses in 10000 scenarios. The results are shown in Table 6.2 in comparison to the ones obtained for the same case study in section 5.2.1. As expected, the maximum MI criterion contributed with slight efficiency improvement for the breakers whose status reconstruction was not impeccable. The difference of measurement selection by MI and the engineering judgment strategy is illustrated in Figure 6.3, with breaker 8 as the representative case. This figure shows once more that the most informative measurements are not necessarily in the adjacent nodes and lines, which is also confirmed by the topology estimation results.

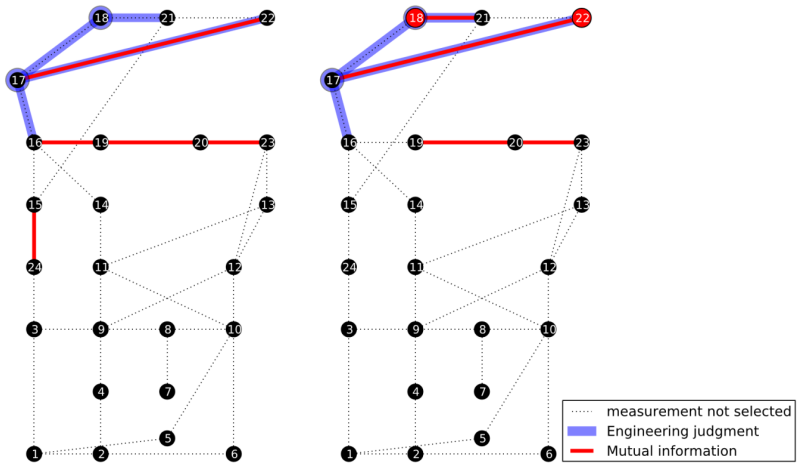


Figure 6.3: Measurements selected for autoencoder inputs with Engineering judgment and Mutual Information criterion (Left: active power, Right: reactive power). This figure aims to indicate how different is measurement selection according to maximum mutual information criterion in comparison to the selection made by engineering common sense.

One could argue that this was not a representative test for the novel measurement selection criterion, since power flows through breakers which contain direct information about breaker statuses, were available. Relevance of the criterion should be tested on a not so trivial case study. Thus, a much more demanding test case from section 5.2.3 is analyzed, where only measurements remote to each breaker are considered as available. This is a type of a stress test: the task is to extract the maximum information from data which are not adjacent to the breakers. The engineering approach, used in Chapter 5, employed measurements from the second order adjacency, i.e. available measurements which are electrically closest to each breaker. Table 6.3 compares these results with the ones obtained using autoencoders defined by the MI criterion. As expected, the benefits from a novel criterion are much more significant for all the breakers in this case study. The proposed MI criterion evidently provides more information (still hidden) to autoencoders then it is contained in the second level of adjacency. Moreover, these results prove once more that most relevant measurements are not necessary also electrically closest to a breaker. It is important to emphasize that any other conventional topology estimator cannot be even run in this case study, since observability is not satisfied in the breaker surrounding.

These results are very significant since they prove the benefits of the mutual information criterion to the particular method, beyond the self-evident theoretical justification. This provides a key upgrade of the proposed topology estimator, since it allows an automatic and efficient autoencoder definition, without the requirement for any user-defined guesses.

Table 6.3: Comparison of engineering judgment and mutual information as criteria for defining autoencoders in a scenario when only measurements remote to the breakers are available. It is significant to note that the MI criterion conveys much more information, resulting in substantially more efficient breaker status reconstruction.

Breaker	**Engineering judgment**			**Mutual Information**		
	Wrong	Correct	Efficiency	Wrong	Correct	Efficiency
1	2232	7768	77.68%	677	9323	93.23%
2	1051	8949	89.49%	552	9448	94.48%
3	1235	8765	87.65%	87	9913	99.13%
4	286	9714	97.14%	142	9858	98.58%
5	2504	7496	74.96%	1155	8845	88.45%
6	0	10000	100.00%	0	10000	100.00%
7	2697	7303	73.03%	2559	7441	74.41%
8	468	9532	95.32%	359	9641	96.41%
9	4449	5551	55.51%	3206	6794	67.94%
10	119	9881	98.81%	94	9906	99.06%
Total	15041	84959	84.96%	8831	91169	90.17%

The relevance of the mutual information criterion in autoencoder performance is investigated in more detail for the two most representative samples: breakers 6 and 9. The idea is to investigate sensitivity in topology estimation efficiency on the amount of MI conveyed on the inputs. The ranking of 10 measurements, sorted in a descending order by their mutual information with the breaker status, is shown in Figures 6.4 and 6.5 for breakers 6 and 9, respectively. As it may already have been noted in the MI maps (Figures 6.1 and 6.2), it is evident that breaker 6 has vastly more impact on the other electric variables than breaker 9. In other words, one can say that it spreads out more information through the available measurements. Particularly, the only significant information sharing from breaker 9 is noted in the active and reactive power flows through it (power flow through the line 19-20).

After calculating the measurement ranking, a sensitivity analysis of topology estimator performance on the value of MI is tested. This is performed in the following way: a set of 15 consecutive measurements, ranked by MI, were selected to serve as input to autoencoders. The first experiment included 15 best ranked measurements. Subsequent 10 experiments were performed with an offset of 1 measurement in a way that the best measurement from the subsequent

experiment set is excluded, and the best one, which is still not considered, is added to the set. For example, in the second test, the first ranked measurement is excluded and the 16^{th} measurement is included. Therefore, the autoencoder is always of the same size, only the window of included measurements over the ranking scale is moved successively towards lower-ranked, progressively excluding the best ranked measurement, while including a one more measurement with lower ranking.

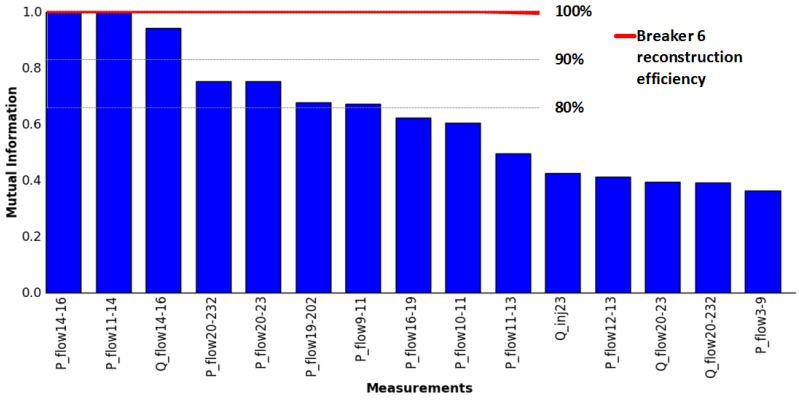


Figure 6.4: Ranking of measurements according to MI criterion regarding breaker 6. The red line represents efficiency of breaker 6 status reconstruction through the described sensitivity analysis. It is of key interest to observe that reconstruction efficiency is kept at 100% despite the 10 most relevant measurements about breaker 6 are considered as unavailable.

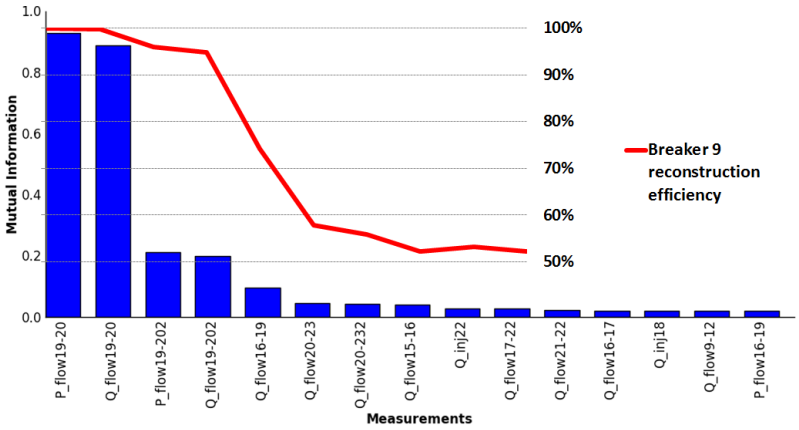


Figure 6.5: Ranking of measurements according to MI criterion regarding breaker 9. The red line represents efficiency of breaker 9 status reconstruction through the described sensitivity analysis. It is significant to note that the efficiency abruptly decreases after the first two measurements are assumed to be unavailable. It is significant that topology reconstruction efficiency is strongly dependent on the amount of mutual information provided by inputs.

The results of breaker status identification efficiency for all the 11 experiments is depicted in a red line in Figures 6.4 and 6.5 for the breakers 6 and 9, respectively. It is important to deduce that breaker reconstruction efficiency is strongly related to the amount of MI. For example, in case of

breaker 9, one can observe that the exclusion of the 2 best-ranked measurements severely hampers the efficiency of the breaker status reconstruction. This is in accordance with the MI ranking which reveal that the 3^{rd}-ranked measurement already contains very poor information about the status of breaker 9. On the other hand, the autoencoders associated to breaker 6 (Figure 6.4) manage to preserve an impeccable performance even without all the 10 best-ranked measurements. If the MI ranking of breaker 6 is observed, this result is easy justifiable due to a large spread of information about its status throughout the available variables.

The most important conclusion derived from these results is: the autoencoder performance is strongly dependent on the amount of mutual information provided by the inputs. This implies that mutual information is a very suitable criterion for definition of local autoencoders in the proposed auto-associative competitive topology estimator.

6.2.2 Case Study - a Substation Topology

The proposed measurement selection criterion based on mutual information is tested here on the split bus problem, i.e. the estimation of substation topology (section 5.3), defined by complex arrangements of circuit breakers. The calculation of MI is the same as for a single breaker, only instead of the 2 possible statuses ($0 = open$; $1 = closed$), the topology variable associated to a more complex topology cell takes multiple discrete values.

For the purpose of establishing a comparative study and testing the benefits of the MI criterion, the same case studies are used as in section 5.3. All the method parameters remain the same, only instead of engineering judgment, MI is considered for defining measurements of local autoencoders.

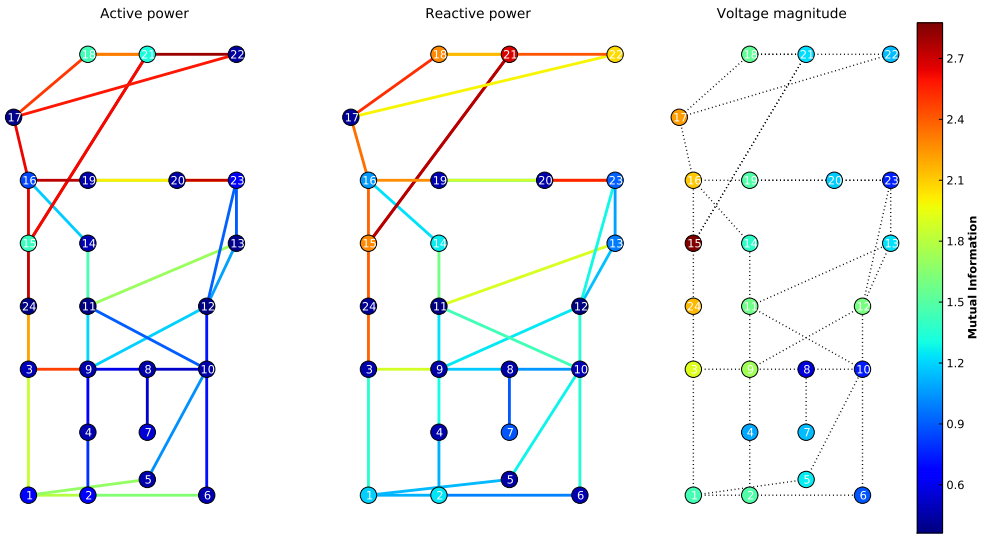


Figure 6.6: Mutual information maps for topology of substation 15. Left: active power. Middle: reactive power. Right: Voltage magnitudes. It is important to note that measurements with significant mutual information are not necessarily adjacent to substation 15.

The Mutual Information between a topology variable and all available measurements evaluates, in fact, the level of observability about a particular topology. Accordingly, one can say Figure 6.6 displays a map of observability for the inner topology of substation 15 (figure 5.12), which can take 35 topologies (possible connection modes) that can be distinguished from external measurements. Three maps correspond to active power, reactive power and voltage measurements, respectively. It is interesting to observe that measurements with high mutual information are not at all necessarily incident with node 15. By knowing all the system parameters, it is interesting to observe these results from the electrotechnical point of view. Specifically, it is evident that power injections which react much more to a topology change are in the P-V nodes, i.e. nodes which contain generator units. On the other hand, in the P-Q nodes, where the injection corresponds to the load rate, mutual information on topology is negligible since the system load changes independently of the topology. Also, as expected, the most relevant information about topology in substation 15 is contained in the voltage magnitude of bus 15.

It can be observed that in the case of a single breaker, mutual information was never beyond 1 and now it may reach major values. This can be explained by the fact that MI is, by definition, upper-bounded by the minimum value of entropy of the two considered variables. For a single breaker, the associated topology variable is binary (open or closed), so its maximum entropy is equal to 1. In the particular case of substation 15, topology variable can take values of 1 to 35, so its estimated entropy is approximately equal to 4.5. Accordingly, the maximum value that mutual information can theoretically contain about topology of substation 15 is equal to 4.5. However, since there is no such a highly informative measurement, the maximum value of MI in this particular case equals to approximately 2.8.

Table 6.4: Test of mutual information criterion for substation topology estimation with external measurements configuration. For all substations, autoencoders defined by mutual information managed to perform at least as efficiently as the one based on engineering judgment.

	Engineering Judgment			**Mutual information**		
Substation	Wrong	Correct	**Efficiency**	Wrong	Correct	**Efficiency**
SE 1	0	10000	**100.00%**	0	10000	**100.00%**
SE 2	0	10000	**100.00%**	0	10000	**100.00%**
SE 9	24	9976	**99.76%**	19	9976	**99.76%**
SE 10	3	9997	**99.97%**	1	9999	**99.99%**
SE 13	0	10000	**100.00%**1	0	10000	**100.00%**
SE 15	14	9986	**99.86%**	10	9990	**99.90%**
SE 16	0	10000	**100.00%**	0	10000	**100.00%**

According to mutual information that each available measurement contains about the substation topology, measurements are being selected for the inputs of competitive autoencoders. Such input selection is expected to be more fruitful in any case other than the one that follows

a criterion of adjacency, the engineering judgment. This hypothesis is confirmed with the results from Table 6.4 which compares these two approaches, considering only external measurements. For all the observed substations, the MI criterion enabled a more efficient topology estimation, if it was below 100% efficient.

Table 6.5: Test of mutual information criterion for substation topology estimation considering internal measurements. For all substations, autoencoders defined by mutual information manage to perform at least as efficiently as the ones defined based on engineering judgment.

		Engineering Judgment			**Mutual Information**		
Substation	Nº AANNs	Wrong	Correct	**Efficiency**	Wrong	Correct	**Efficiency**
SE 1	42	899	9101	**91.01%**	86	9914	**99.14%**
SE 2	42	847	9153	**91.53%**	133	9867	**98.67%**
SE 9	50	3669	6331	**63.31%**	2251	7749	**77.49%**
SE 10	50	2849	7151	**71.51%**	1415	8585	**85.85%**
SE 13	42	2921	7079	**70.79%**	517	9483	**94.83%**
SE 15	42	1667	8333	**83.31%**	654	9346	**93.46%**
SE 16	50	2556	7444	**74.44%**	1106	8894	**88.94%**

The same experiment is repeated for the case which considers power flow meters internal to substations. The results from Table 6.5 indicate that the MI criterion contributes to an enormous improvement of efficiency in topology estimation for all the substation samples.

All results reported in this section justify MI as a proper criterion for the definition of proposed auto-associative topology estimator. In other words, the selection of electric variables which contain a maximum MI with topology variables results in a substantial improvement of topology estimation efficiency. Moreover, it offers a systematic and automated way for defining the mosaic of local autoencoders without any user-defined parameters. The following section proposes a methodology for solving another key important concern - arrangement of the topology cells.

6.3 Mutual Information for Optimal Topology Cell Arrangement

The maximum mutual information criterion provides a mechanism for an automatic definition of most informative inputs regarding the topology of a group of circuit breakers, i.e. a topology cell (TC). In respect to the proposed competitive autoencoders scheme, a number of required autoencoders will be directly defined by configuration of TC, i.e. by a number of topologies it could take. In the previous chapter, it was demonstrated that the definition of topology cells in a substation had a drastic impact on a number of autoencoders needed to be trained, as well as on the final estimation results. However, until now, TC arrangement was defined by a guess and this section intends to provide a proper methodology for the resolution of this problem.

In a decentralized local concept of local autoencoders mosaic, the most elementary TC can be a single breaker. Alternatively, in a global scheme, a single TC corresponds to the entire power system. In reality, an optimal solution is expected to be found somewhere in between, i.e. topology cell may be attributed to a substation, or to some specific groups of circuit breakers. Since all these solutions evidently lack some theoretical justification, the intention is to adopt a criterion for optimal topology cell arrangement in order to provide high performances in regard to computation requirements and estimation efficiency.

For example, in a substation with n circuit breakers, the maximum number of configurations is theoretically 2^n, provided that breaker statuses change independently. If TC includes all the breakers, 2^n autoencoders are required. Alternatively, if there are n autoencoders for n breakers, only $2 \times n$ autoencoders are sufficient.

The solution for topology cell arrangement is offered again from an information theoretic point of view. The main idea is the following: the topology variables which contain a significant information flow between each other need to be grouped in the same topology cell. For example, lets assume a trivial case of 2 breakers which always take opposite statuses, i.e. two topologies are possible: $(0,1)$ or $(1,0)$. If one TC is associated to each breaker, 4 autoencoders will need to be trained. On the other hand, if two TC are merged in one, just 2 autoencoders are required. Hence, the principal intention is to minimize number of considerable topology alternatives. Thus, such optimization of TC arrangement is expected to contribute with a double benefit: improve topology estimation efficiency and reduce computational and memory requirements.

Therefore, the idea is to group breakers whose statuses are, at a certain extent, dependent into the same topology cell. A logical choice for measuring of such dependency between the discrete topology variables is the Mutual Information. Let T_1 and T_2 represent two discrete topology variables associated to topology cells TC_1 and TC_2, respectively. Renyi's Quadratic Mutual Information for the two topology cells T_1 and T_2 is then calculated as follows:

$$I(T_1,T_2) = \log \sum_j \sum_i \frac{p(t_{1i},t_{2j})^2}{p(t_{1i})p(t_{2j})} \tag{6.11}$$

The proposed strategy is to define a TC arrangement on such a way that MI between them is minimal. As it was previously discussed, the calculation of multivariate mutual information is computationally very demanding, so a greedy algorithm is proposed (algorithm 3), which considers only MI for pairs of variables. The idea is to initiate the TC arrangement on the way that each topology cell corresponds to a single breaker. In the subsequent steps, TC couples which contain significant mutual information are being merged to a single TC. This process is iteratively continued until there is no any significant mutual information value between the newly formed topology cells. This way, the number of topology alternatives is expected to be minimized and accordingly computational and memory requirements proportional to number of autoencoders.

For the purpose of clarity, the proposed algorithm for optimum TC arrangement is demonstrated on the sample of substation 1. A scheme of circuit breakers for the substation 1 is presented in Figure 6.8. If one assumes that the status of each circuit breaker is defined completely

Algorithm 3 Optimization of topology cells arrangement by the mutual information criterion.

1: Initialize arrangement of TC: *1 breaker = 1 TC*
2: Calculate MI between all the pairs of TC variables
3: **while** any TC couple contains some considerable MI value **do**
4: Merge TC with considerable MI into a single TC
5: Calculate MI between all the recent TC variables
6: END:

independently to the others, the MI between all the TC will be zero. In that scenario, the TC arrangement optimization algorithm will end just after the initialization and the proposed solution will correspond to the largest possible number of topology cells (as in Figure 5.13). Accordingly, for the $2^{11} = 2048$ possible configurations of 11 breakers, only 42 autoencoders will be required.

However, in this particular case study, only 136 distinguishable topologies of substation 1 are considered. Figure 6.7 indicates MI between all the possible couples of topology cells, separately for the cases when a bus-tie breaker is open and closed. After the first iteration, the following TC pairs are merged: $\{2,3\}$, $\{4,5\}$, $\{6,7\}$, $\{8,9\}$, $\{10,11\}$, $\{12,13\}$, $\{14,15\}$, $\{16,17\}$, $\{18,19\}$, $\{20,21\}$. The final arrangement results with 11 topology cells as illustrated in Figure 6.8. The cases of closed bus tie breaker (no split bus) are reduced to a single-bus scenario and each incident line can be connected or disconnected to the bus. When the bus-tie breaker is open, three autoencoders are required for each possible line connection scenario: connected to bus 1, connected to bus 1' or disconnected. This TC arrangement requires finally only 27 autoencoders for estimation of topology in substation 1. One can observe that any further merging or unmerging will increase the number of possible topology alternatives. In other words, the TC arrangement, suggested by the proposed algorithm, should correspond to the minimal number of autoencoders.

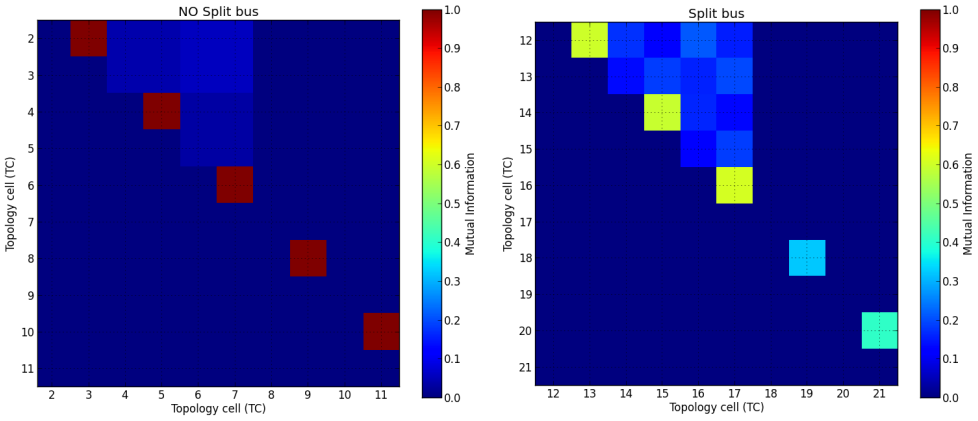


Figure 6.7: Mutual information between topology cells for the initial arrangement in substation 1. Left: bus-tie breaker closed - no split bus. Right: bus-tie breaker open - split bus.

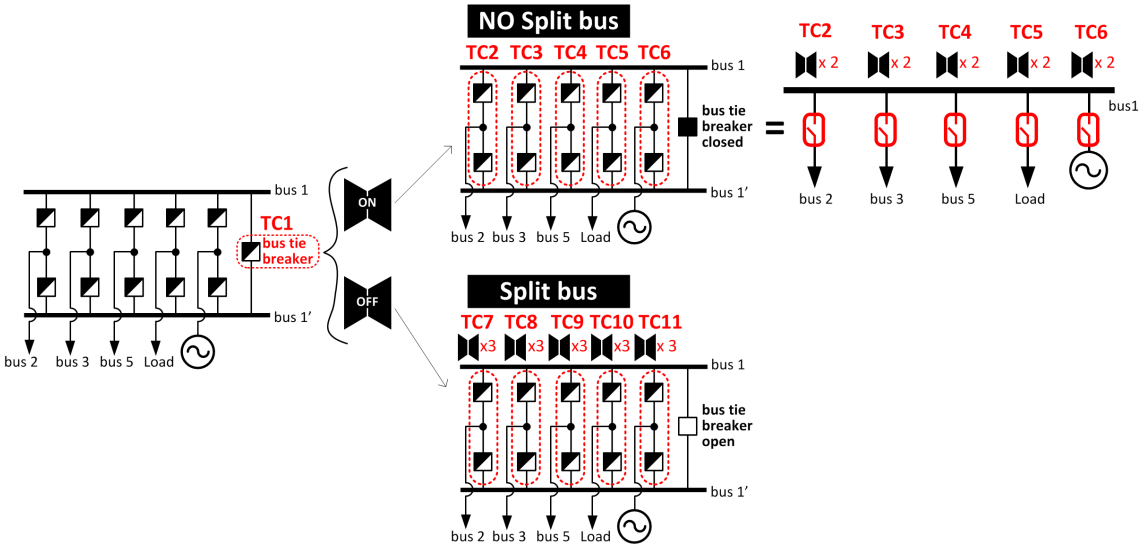


Figure 6.8: Optimized topology cell arrangement and scheme of competitive autoencoders in substation 1. With 11 topology cells, a minimal number of autoencoders is achieved.

Table 6.6: Results of substation topology estimation with initial topology cells arrangement and after the arrangement is optimization. The proposed algorithm for optimization of TC arrangement contributes with improvement in topology estimation efficiency reduces number of required autoencoders.

			before TC optimization		**after TC optimization**	
Substation	Nº breakers	Nº topologies	Nº AANNs	**Efficiency**	Nº AANNs	**Efficiency**
SE 1	11	136	42	**99.14%**	27	**99.73%**
SE 2	11	136	42	**98.67%**	27	**99.76%**
SE 9	13	602	50	**77.49%**	32	**79.14%**
SE 10	13	602	50	**85.85%**	32	**85.96%**
SE 13	11	136	42	**94.83%**	27	**95.30%**
SE 15	11	136	42	**93.46%**	27	**97.28%**
SE 16	13	288	50	**88.94%**	32	**94.54%**

The proposed algorithm for TC arrangement is tested for all the observed substations. For the purpose of comparability, two TC arrangements: initial arrangement with a largest possible number of topology cells and the arrangement proposed by the MI optimization procedure. The results in Table 6.6 indicate a significant efficiency gain after the topology cells are optimally arranged according to the proposed MI criterion. Detailed results for each topology cell of

Table 6.7: Detailed topology estimation results for substation 1, using optimized arrangement of topology cells.

	TC	Wrong	Correct	Efficiency
Bus-tie breaker	1	0	10000	100.00%
No split bus	2	0	5641	100.00%
	3	0	5641	100.00%
	4	0	5641	100.00%
	5	0	5641	100.00%
	6	0	5641	100.00%
Total: NO Split bus		0	5641	100.00%
Split bus	7	0	4359	100.00%
	8	12	4347	99.72%
	9	11	4348	98.72%
	10	0	4359	100.00%
	11	4	4355	99.91%
Total: Split bus		27	4332	99.38%
TOTAL		27	9973	**99.73%**

substation 1 are represented in Table 6.7.

The proposed method for optimization of topology cell arrangement offers a significant upgrade for the novel auto-associative topology estimator from Chapter 5 in regard to the following benefits:

- ◇ Reduction of computational and memory requirements, due to minimization of the required number of autoencoders.
- ◇ Provides an automatic settings for the topology estimator, eliminating needs for some user-defined guesses.
- ◇ Enhancement of the topology estimation efficiency.

6.4 Mutual Information - a Measure of Statistical Observability

The mutual information is proposed in this chapter as a criterion for selecting measurements which are most relevant for consideration of the auto-associative topology estimator. One can also say, from an information theoretic point of view, that mutual information quantifies observability of the system topology. Since mutual information is a general measure of information between variables, this concept may be extended to any other electric variable. Hence, if one wants to improve the

estimation of some electric variable, a mutual information map will directly indicate the locations from which most of the information can be retrieved. This is another key issue within the proposed paradigm of approaching state estimation from an information theoretic point of view. The novel concept of perceiving observability in power systems is called ***Statistical Observability*** (**SO**).

In comparison to the classical observability analysis, the statistical observability is essentially different in its nature and is not proposed as an alternative, but more as a supplement. The classical observability analysis has a task to determine the possibility of finding a unique solution (state estimate) of a system of equations, regarding the available set of measurements. Thus, it gives an answer to the question if a system is observable or not, considering the type and location of the available measurements, as well as topology of the system. This concept of observability is binary and allows definition of function O, as a way of mapping the available measurements onto $\{0,1\}$. This is an outcome-focused definition: when a set of measurements M is such that a single state vector can be estimated, $O(M) = 1$ and the system is observable. Alternatively, if $O(M) = 0$, the system is not fully observable and the analysis proceeds with a definition of observable islands and non-observable parts of the system.

The proposed concept of Statistical Observability resides on completely different paradigms, borrowed from the Information Theory. For example, the SO of a variable is not dependent on the analytical solution of a system of equations, but on the information flow among the electrical variables, calculated from the available dataset. Moreover, it maps to $[0,1]$ (instead of $\{0,1\}$) expressing the amount of information provided about a particular variable of interest. It also measures a level of impact one variable has on the other, given the available dataset, so it is also very related to sensitivity analysis. Unlike the classical analysis of sensitivity, the SO captures more than just electric rules: it may also uncover additional information hidden in a historical dataset, like load curves, electric power dispatch, etc.

Similarly to a classical observability analysis, SO is run prior to the estimation process. It allows to systematically identify which variables are most valuable to be considered for estimation of some unknown value. This is important information for the estimation methods proposed in this thesis, which rely on information theoretic descriptors.

In any case, the SO concept can have a broad range of applications in state estimation. For example, it can suggest where some new measurement devices need to be installed in order to enable an efficient reconstruction of a variable of interest. Also, it may be an important upgrade to processes like gross error identification, reconstruction of missing measurements or generation of pseudo-measurements.

In order to demonstrate the relevance of the proposed SO concept, some representative statistical observability maps (or mutual information maps) are analyzed, considering one-to-one MI values with active power, reactive power and voltage magnitudes, respectively. Figure 6.9 illustrates SO maps for the active power injection in bus 21, which effectively corresponds to active power generation in bus 21. It is significant to observe the high spread of information through all the nodes which contain power loads and power generation. This could be explained by the following: the generator from bus 21, due to its operational costs (dataset is generated by

OPF), covers the peak of the electric power dispatch and closely follows the load variation. An insignificant level of MI with P_{inj1} is obtained only in nodes with constant power injection.

Figure 6.10 illustrates another sample of the SO maps for voltage magnitude measurement in bus 3. Since it is a P-Q bus type, it means that its voltage magnitude is not being regulated, i.e. it is just a result of a particular load flow. Now, it must be more clear why a significant mutual information is detected with active and reactive power injections and voltage magnitudes in other P-Q nodes.

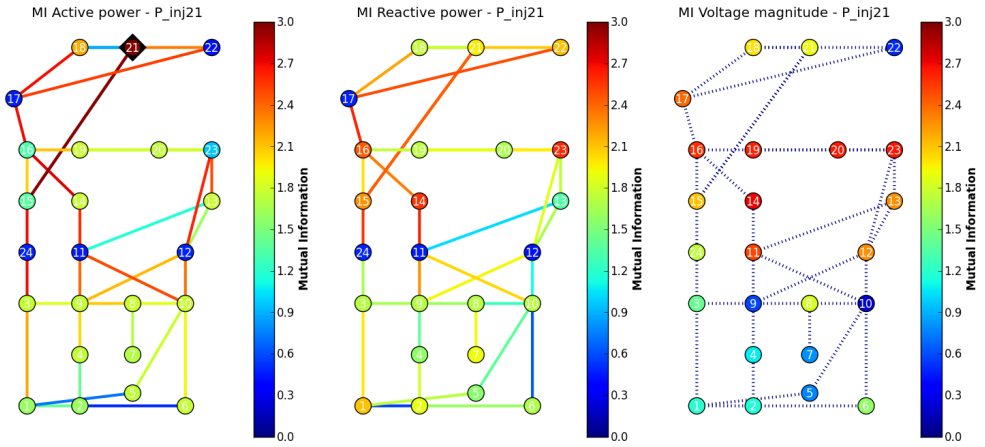


Figure 6.9: Observability map for active power injection in bus 21, in regard to active power (left), reactive power (middle) and voltage magnitudes (right).

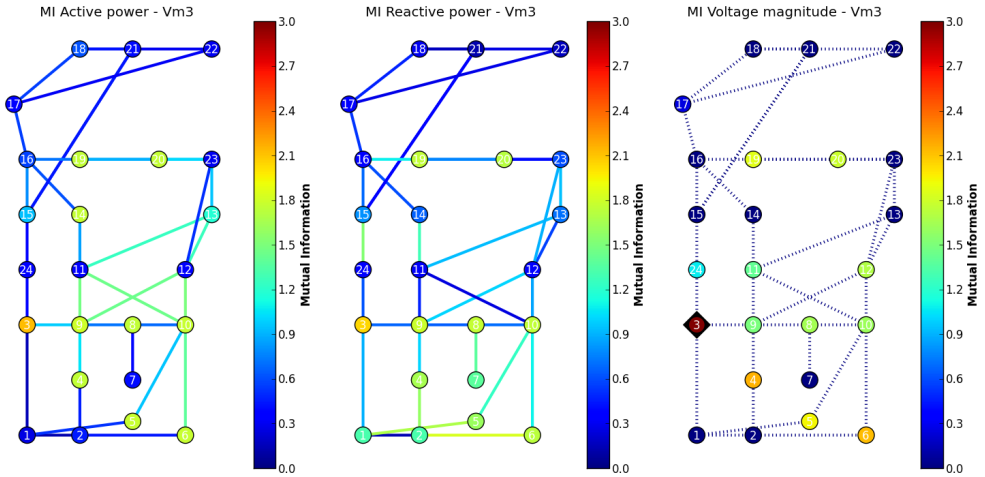


Figure 6.10: Observability map for voltage magnitude measurement in bus 3, in regard to active power (left), reactive power (middle) and voltage magnitudes (right).

In summary, in regard to the traditional perception of observability, the SO cannot be definitely seen as an alternative, but more as a complementary analysis. The classical observability analysis is concerned with sufficiency of the available measurement set for solving the system of equations

that result in estimative of state of the entire system. On the other hand, statistical observability does not regard at all the algebraic requirements for state estimation procedure. It captures information that exists among measurements through the historical dataset and supports system's observability by enabling screening and reconstruction of measurements acquired at the certain moment.

The novel concept of Statistical Observability announces a broad field of applications in electric power systems. However, one cannot disregard some limitations of the concept presented in this chapter. Mutual information, as a main descriptor of statistical observability, was calculated only for pairs of measurements since a straightforward multivariate model requires an unreasonable computational effort. This simplification is to some extent acceptable for the purpose of topology estimation but may not be justifiable in some other, more complex applications. Therefore, it is admissible that some other information theoretic descriptor may be proposed in the future as a more suitable solution. Further research on these issues goes beyond the envisaged contents of this thesis.

6.5 Conclusions - the Statistical Observability

This thesis establishes a novel decentralized methodology, based on a mosaic of local auto-associative neural networks for screening electric variables and estimation of topology. An issue that was left unaddressed before this chapter deals with the optimization of decentralized SE architecture for the purpose of an efficient extraction of maximum information. Specifically, if the problem of topology estimation is observed, a key question asks about the best selection of input electric variables for achieving the most efficient topology estimation in a power system. Logically, the selection of input variables must have an important impact on the performance of autoencoders.

Therefore, one intends to set a criterion in order to automatically select most informative measurements about the system topology. For that purpose, another information theoretic measure is borrowed, called Mutual Information. It quantifies the information that one variable contains about another thus decreasing the uncertainty about it. When focusing on the problem of topology estimation, the idea is to select a group of measurements that contain maximum information about the statuses of a group of breakers.

The mosaic of autoencoders defined in Chapter 5 was based on a hypothesis that a change in a breaker status will impact power flows and node voltages mostly in a local area. In order to theoretically justify this assumption based on engineering judgment and allow a systematic definition of the decentralized model, the maximum mutual information criterion is proposed. The result from the experiments have shown that the intensity of impact of topological change is not necessarily limited to adjacent nodes and lines, but strongly depends on the amount of MI provided to the inputs.

The most valuable indicator about the relevance of the criterion is the significant improvement in topology estimation efficiency when the engineering judgment criterion is substituted by the

maximum mutual information criterion. Not only that it evidently allows building better intelligent systems, but it also offers a strong theoretical background and eliminates the need for any user-defined guesses.

For auto-associative topology estimator proposed in Chapter 5, the maximum mutual information criterion offers a mechanism for an automatic definition of an optimal scheme of competitive local autoencoders. According to the mutual information between all the topological variables and between the analog measurements and the topological variables, one can optimize the number of autoencoders and their borders in regard to topology estimation efficiency. This concept offers a significant support to all the methodologies proposed in this thesis.

Furthermore, even larger capabilities need to be emphasized, if considering possibilities for extension of the concept to any type of electric variable. Generally, MI provides information that can be used to reconstruct a corrupted variable. In that sense, a new general paradigm called Statistical Observability is established. Unlike the classical definition of observability, this approach is not dependent on the analytical solution of a system of equations, but rather on informational content of a joint probability distribution resulting from the solution of power flow equations. Although it is not a binary concept (observable/unobservable), it provides quantification in $[0, 1]$ about how much of it is observable. In any case, the statistical observability cannot be envisaged as an alternative for the classical observability analysis. It must be understood as a tool that supports the observability assessment and improves redundancy by enabling the generation of pseudo-measurements or the reconstruction of any missing electric variable.

In conclusion, the Statistical Observability concept, in a certain way, closes the main paradigm established in this thesis that has a key role in perceiving the power system topology from an information theoretic point of view. It also paves the way towards many other applications, not necessarily only in the area of power system analysis. Future work on this concept should be aimed at the extension of the concept to multivariate cases, and is expected to bring yet more improvement in quality for the most challenging applications.

Chapter 7

Conclusions

7.1 General Conclusions

The main goal of this thesis is to establish new paradigms for state estimation, in line with the modern trends in evolution of power systems. Nowadays, we are witnessing some fundamental operational changes that power systems are undergoing, mostly due to the following: deregulation of the electricity market; activation of distribution grids by a rapidly growing penetration of distributed generation; growth of the installed power in stochastic energy sources on all voltage levels; development of infrastructure ready to accept electric vehicles; integration of modern power system operational concepts like micro-grids or smart-grids. It is evident that all these issues will have severe impacts on the operation and control of the power system on all voltage levels. The power system state estimator, as one of core functions of EMS and DMS, also needs to be subjected to some conceptual changes in order to be suitable for the new environment. The novel challenges that a modern state estimator is expected to meet are the following: the extension of state estimation to distribution grids; the increasing requirements for distribution system monitoring; the growing implementation of PMU measurements; the increased requirements of accuracy; the shorter refreshing time period about the system state. The conventional static state estimation models, which solve a global system of equations through certain function of residuals, cannot be considered as a satisfactory solution for meeting these requirements.

Motivated by all the aforementioned issues, this thesis proposes a set of novel and unconventional concepts which are used to build a state estimator with high level of robustness, reliability and flexibility. The novel state estimator follows a paradigm different from the classical approach currently followed by the industry: instead of a single global state estimator, this thesis proposes a set of semi-autonomous local estimators that rely on local information and that are operated using concepts from Information Theory rather than just applying the classical Kirchhoff Laws.

Therefore, the metrics and descriptors incorporated in the methods and from which the entire valuable features are inherited, are all borrowed from Information Theory and Information Theoretic Learning. Because of this, one can say that this thesis intends to shed a new light on the

state estimation problem by approaching it from information theoretic point of view. Moreover, this thesis establishes ways to move from traditional global concepts, defined as mathematical programing problems, towards decentralized and localized models that perceive state estimation as a problem dealing with information. Therefore, the main hypothesis set as a foundation of this work assumes that information theoretic paradigms may offer a fruitful background for the establishment of a modern state estimator, able to meet the most demanding relevant challenges.

The state estimator is a complex system that produces an image of reality based on a perception gathered by sensors and translated into measurements, and whose values are inputs to the SCADA at the control center. The state estimation in a system without noise is fairly simple. The traditional approaches consider the presence of noise in data with a relatively small amplitude, and this assumption allowed the estimation systems with reasonable accuracy to be built.

However, when the signal-to-noise ratio decreases, it is known that the estimation becomes less reliable. In power systems, the presence of a large noise component is usually translated into the sentence "presence of gross errors". This thesis has addressed the issue of a measurement set contamination by gross errors. Furthermore, the concept of gross error is extended beyond the magnitude evaluation in two additional ways: a) the absence of measurements (when supposed to be present); b) the errors in topology information.

A first general conclusion from the thesis work presented may be summarized in the following way: the adoption of Information Theoretic Learning concepts allowed one to address the diversity of aspects related to gross errors, in the broad definition referred to above, and a unified perspective with practical results. The following sections will address the major contributions of the thesis in depth, highlighting the main conclusions in each of the main topics addressed.

7.2 A New Hybrid State Estimator Based on Correntropy

The cornerstone of this thesis is a paradigm adopted from the Information Theoretic Learning: a framework that enables the estimation of information theoretic descriptors directly from the data in order to use them in the adaptation of linear or nonlinear filters and machine learning applications. A particular interest in this thesis is focused on the Correntropy function - a novel measure of similarity, whose robust features are well-recognized in machine learning and signal processing. It is for the first time adopted in the power system analysis throughout this thesis. The criterion of maximum Correntropy enables a powerful fusion of a straightforward, fast and robust static state estimator with a decentralized mosaic of local data filters that contribute to robustness and reliability of the method without excessive sacrificing of precious computational resources.

After decades of research attempts to improve robustness in state estimation, it can be generally concluded that robustness in bad data processing is gained at the cost of computational efficiency. In other words, a high level of estimation robustness usually leads to a lack of feasibility in real time environment. **The proposed hybrid estimator, based on Correntropy (Maximum Correntropy Criterion) and auto-associative local data screening, has managed to overcome this trade-off thanks to its decentralized and localized architecture, i.e. it offered extremely**

high robustness scalable even to large-scale systems. The benefits of this method become particularly evident in cases when successful processing of bad data is seriously threatened by poor redundancy or even lack of system observability. Unlike the conventional estimators, the proposed method is able to identify and correct gross errors even in critical measurements. In regard to other known methods, it is important to emphasize, that the novel estimator does not perform worse in any state estimation segment. More specifically, when measurements are free of gross errors, it will behave in the same way as the least squares estimator. Moreover, in case of a single gross error or multiple non-interactive gross errors, the MCC estimator will naturally detect, identify and eliminate their harmful impact on the final solution. Finally, the gross errors scenarios, which are the toughest to be identified by most of the traditional estimators, will be treated with engagement of specific data screening, also based on information theoretic criteria. Computational requirements of this processing is efficiently limited, independent on the system size, by focusing it only on the narrow local corrupted subsystem. Therefore, the proposed hybrid method avails the best features from its components, by a selective adaptation according to severity of a specific problem.

Furthermore, the MCC criterion allowed the development of a new diagnosis tool that signals the "healthy state" of data by confirming the solution did not change through the process of kernel annealing. This forms a basis for a *detection* phase in the state estimation process, which guarantees that only suspect data sets are sent for further analysis through the subsequent phases of *identification* and *quantification*. The function of gross error quantification is selectively activated through a decentralized measurements screening with mosaic of local autoencoders. This results with a substantial stride in robustness enhancement, which does not threaten real time feasibility. The adoption of MCC criterion became, therefore, not only an algorithmic trick but a fully based theoretical tool, in its own right.

7.3 Decentralized Topology Estimator

The information theoretic perspective on power system state estimation was extended to a problem of erroneous or incomplete topology information. The same measurement pre-processing methodology is used here in order to capture the information between the electric and the binary topological variables associated to the circuit breaker statuses. This work showed that **a decentralized architecture, based on competitive, auto-associative local estimators, allows a remarkably efficient reconstruction of entire topologies, with computational investments that are practically negligible**. The benefits of the method, in regard to the existing conventional solutions, are especially relevant in occasions when only measurements external to a substation are available, or when the redundancy level is poor in a subnetwork of interest.

A key procedure of the proposed topology estimator is performed offline, when the architecture is being built according to the available historical dataset. In order to address this issue, an important paradigm is again borrowed from the Information Theory, which allows an automatic definition of the architecture aiming to extract a maximum of information from available data.

The criterion of Maximum Mutual Information allowed an efficient decentralization, i.e. definition of the borders of the most relevant local areas. This enabled a more efficient intelligent system, in the sense to satisfy the trade-off between efficiency and consumption of computational and memory resources.

The properties of adopted mutual information criterion are used to establish foundations of a more valuable general concept called **Statistical Observability**. All concepts proposed in this thesis are relying on some amount of extractable information, among any type of electric variables, weather measurements of power and voltage, topological variables or network parameters. **Unlike the classical definition of observability, that is concerned with the analytical solution of a system of equations, the Statistical Observability concept relies on information content of joint probability distributions resulting from the available dataset.** It is important to emphasize that the novel concept does not offer an alternative, but more a supplemental tool that is able to convey a better starting point for a fast and efficient power system state estimation.

7.4 Information Theoretic Power System Analysis

All the concepts adopted by this thesis for the state estimation purposes may be characterized as an attempt to approach state estimation functions from a perspective of information theory. Furthermore, they all have one mission in common: the establishment of framework for a decentralized and distributed system analysis, which is considered of key importance in regard to modern trends in power systems: micro-grids, smart grids, intelligent agents, etc. All the key contributions this thesis brings to the theory of power system state estimation are systematically resumed hereafter:

- ▷ The adoption of Information Theoretic paradigms in state estimation opens the way towards some new perspectives in power system analysis.
- ▷ A novel state estimator based on Maximum Correntropy Criterion.
- ▷ A decentralized measurements screening built on mosaic of local autoencoders, evolutionary particle swarm optimization (EPSO) and the Correntropy function.
- ▷ A novel, fast and robust hybrid state estimator which couples the former two concepts.
- ▷ A novel decentralized topology estimator, based on competitive auto-associative neural networks. It is characterized by negligible computation requirements and high robustness regarding redundancy level and gross errors.
- ▷ The establishment of statistical observability concept, which forms a new stepping stone in definition of optimal architectures for an efficient decentralized estimator.
- ▷ The proposal of decentralized local state estimation architecture, which may represent a substantial step towards the integration with the modern power system concepts as micro-grids and smart-grids.

7.5 Publication of Results

The work addressed in this thesis was reported to the scientific community in scientific articles published in most prestigious international journals and by presentation at conferences. In Appendix D a list of publications that resulted from the work described can be found.

7.6 Discussion about Future Work Perspectives

The main scope of this thesis was to establish new paths in power system state estimation, that intend to meet the challenges that modern power systems are coping with. However, a number of issues was left unaddressed. The identification of some of these aspects forms the basis of our future work proposal. Among many aspects that could be discussed, the following visions of possible and needed future developments should be emphasized:

– All the proposed methods were tested on system samples that are more identifiable with transmission grids. Although they are also conceptualized for implementation in distribution state estimation, special aspects that may emerge in such cases have not been presented in particular detail. Therefore, future engagement related to these paradigms would be to adapt or to test the developed methods specifically for distribution networks or even a low voltage environment.

– We are witnessing an increasing penetration of PMU measurements that will incontestably cause a revolution in the monitoring and state estimation of a power system. Although they are still not installed to a considerable extent, it is reasonable to expect soon a need for a hybrid estimator that will process in parallel both SCADA and PMU measurements. Moreover, in a perhaps not so distant future, when the PMUs reach observability and redundancy by themselves, the relevance of traditional SCADA measurements will be challenged. With PMUs only, state estimators will in fact turn to be linear and much simpler. The research reported in this thesis does not specifically address these scenarios. However, all the proposed methodologies have proved to be suitable for processing any type of electric variables. Accordingly, it would be significant to specifically investigate capabilities of the methods proposed in this thesis for processing of PMU measurements and their fusion with the SCADA measurements.

– The concept of statistical observability, established in Chapter 6, opens a new perspective for quantification of interaction among electric variables and provides key support in efficiency and automation of the proposed data processing. However, mutual information, in the form adopted, shows insuperable computational limitations in multivariate cases. Due to these reasons, the thesis proposed the use of a simplified greedy algorithm, which processes measurements one by one when forming the most relevant group. Alternatively, it is well known that the set of n best individual measurements does not necessary form the best vector of n measurements. This is considered to be an important research topic to extend the mutual

information concept to multivariate contexts or to propose some other descriptor that will be able to take into account multivariate cases. This is expected to significantly contribute to improvement of the proposed decentralized model.

This thesis paves the way towards some new perspectives established on a strong theoretical justification: the application of Information Theory concepts. This can be considered a novelty in theory of state estimation in power systems. All the methods are built with strong decentralized and localized architecture, envisaged for compatibility with modern power system concepts. Accordingly, future applications of this work in many aspects of power system analysis are expected to be rooted in the methodologies proposed in this thesis.

Appendix A

Network Test Cases and Database Generation

This Appendix provides details about the electric power networks used as test case studies in this thesis. It also describes the methodology for generation of measurement set for the power system state estimation. It mostly contains active/reactive power flows, active/reactive power injections and voltage magnitude measurements.

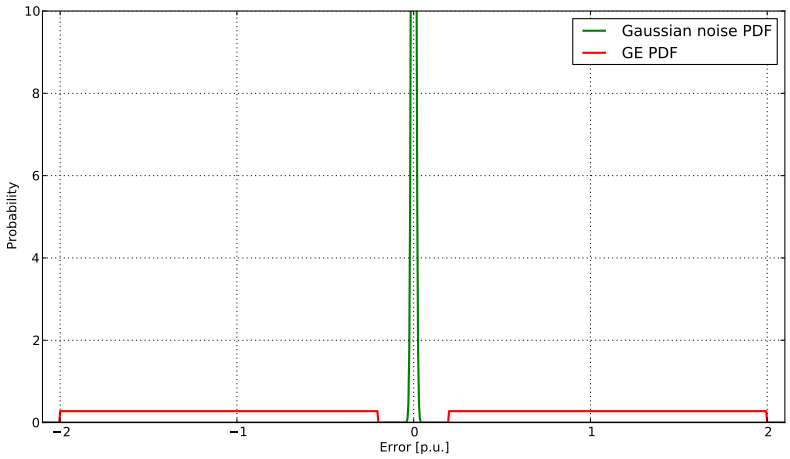


Figure A.1: PDF of gross errors in relation to Gaussian-distributed noise with standard distribution of $\sigma = 0.01 p.u.$. GE are uniformly distributed in the range of 20-200 standard deviations of the normal errors.

Most of the case studies in this thesis include some GE in measurements dataset in order to test the robustness of the estimators. A gross error is defined by a value of 20-200 standard deviations of the Gaussian-distributed errors. Therefore, GE are generated according to the probability density function, as presented in Figure A.1. The green line determines assumed Gaussian distribution of noise, i.e. PDF of expected errors of measurements readouts. The red line represents uniform distribution of GE. After GE are generated according to the assumed PDF, they are added to a randomly selected measurement.

A.1 4-bus Test System

A.1.1 Network Parameters

Testing of the MCC state estimator from Chapter 3 and the hybrid auto-associative MCC state estimator from Chapter 4 was performed on the 4-bus AC system from [128]. Figure A.2 presents single-line diagram of the system. Locations of the power flow meters were considered differently for various levels of measurements redundancy. Characteristics of the buses and parameters of the power lines are given in Tables A.1 and A.2, respectively.

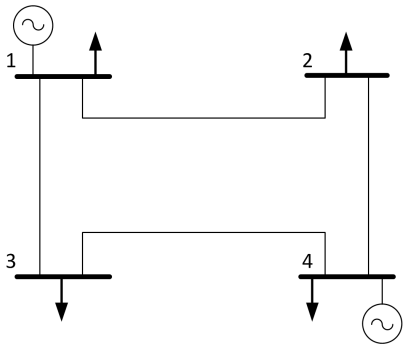


Figure A.2: Single-line diagram of the 4-bus system.

Table A.1: Buses characteristics of the 4-bus test system.

bus	type	V_n [kV]	P_D [MW]	Q_D [MVAr]	P_G [MW]	Q_{Gmin} [MVAr]	Q_{Gmax} [MVAr]	B_s [MVAr at V = 1.0 p.u.]	V_{sp} [p.u.]
1	REF	230	50	30.99	0	-100	100	0	1.02
2	PQ	230	170	105.35	-	-	-	0	-
3	PQ	230	200	123.94	-	-	-	0	-
4	PV	230	80	49.58	318	-100	100	0	1

Table A.2: Branches characteristics of the 4-bus test system.

Branch $i-j$		R	X	B_{sh}	a
bus i	bus j	[p.u.]	[p.u.]	[p.u.]	
1	2	0.01008	0.0504	0.1025	-
1	3	0.00744	0.0372	0.0775	-
2	4	0.00744	0.0372	0.0775	-
3	4	0.01272	0.0636	0.1275	-

A.1.2 Measurements Generation

Each measurement scenario is synthetically generated from the solution of a converged power flow. Afterwards, Gaussian-distributed random errors were added to the power flow solution in order to mimic the real measurements. Hence, measurements generation procedure includes the following steps:

- Load level generated randomly with uniform distribution in the range $[30\%, 100\%]$.
- Load on each bus is randomly modified from the base case within the range of $\pm 10\%$.
- Zero-mean Gaussian errors are added to the power flow solution, with standard deviation of $\sigma = 0.005 p.u.$ (1 p.u. corresponds to 100 MVA) for both power and voltage magnitude measurements.
- If considering the gross errors, they were generated uniformly with the magnitude of 20-200 standard deviations of Gaussian errors and added to a randomly selected measurement.

A.2 IEEE 24-bus Test System

A.2.1 Network Parameters

The network sample with 24 buses is based on the IEEE RTS 24-bus system from [130], with slight modifications. Single-line diagram of the system is illustrated in Figure A.3. All network parameters required to run a power flow calculation of the 24-bus system are given in Tables A.3 and A.4.

Table A.3 contains the following information relevant to the buses:

bus type
- PQ active and reactive injections are known and fixed
- PV active power is known, while generator regulates reactive power in order to reach the specified point voltage V_{sp}
- REF reference node

V_n voltage level

V_n voltage level

P_D active power demand

Q_D reactive power demand

P_G active power generation

Q_{Gmin} minimum reactive power generation

Q_{Gmax} maximum reactive power generation

B_s shunt susceptance

V_{SP} voltage magnitude set-point

Table A.4 contains the following information relevant to network branches:

bus i, j	two end nodes of a branch
R	resistance
X	reactance
B_{sh}	line charging susceptance
a	transformer off nominal turns ratio

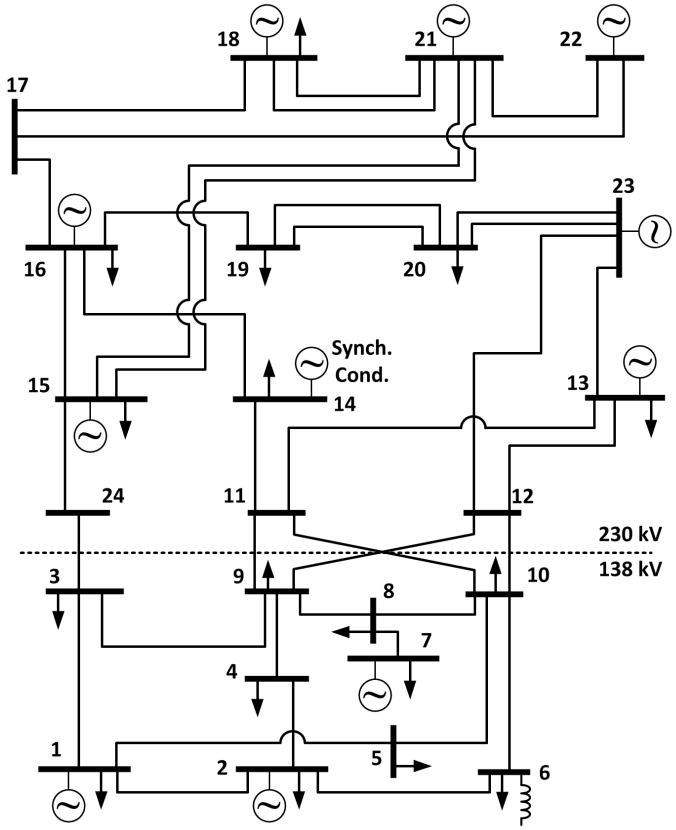


Figure A.3: IEEE RTS 24-bus system [130].

Table A.3: Buses characteristics of the IEEE 24-bus test system.

bus	type	V_n [kV]	P_D [MW]	Q_D [MVAr]	P_G [MW]	Q_{Gmin} [MVAr]	Q_{Gmax} [MVAr]	B_s [MVAr at V = 1.0 p.u.]	V_{sp} [p.u.]
1	PV	138	108	22	172	-50	80	0	1.035
2	PV	138	97	20	172	-50	80	0	1.035
3	PQ	138	180	37	-	-	-	0	-
4	PQ	138	74	15	-	-	-	0	-
5	PQ	138	71	14	-	-	-	0	-
6	PQ	138	136	28	-	-	-	-100	-
7	PV	138	125	25	240	0	180	0	1.025
8	PQ	138	171	35	-	-	-	0	-
9	PQ	138	175	36	-	-	-	0	-
10	PQ	138	195	40	-	-	-	0	-
11	PQ	230	-	-	-	-	-	0	-
12	PQ	230	-	-	-	-	-	0	-
13	REF	230	265	54	285.3	0	240	0	1.02
14	PV	230	194	39	0	-50	200	0	0.98
15	PV	230	317	64	215	-50	110	0	1.014
16	PV	230	100	20	155	-50	80	0	1.017
17	PQ	230	-	-	-	-	-	0	-
18	PV	230	333	68	400	-50	200	0	1.05
19	PQ	230	181	37	-	-	-	0	-
20	PQ	230	128	26	-	-	-	0	-
21	PV	230	-	-	400	-50	200	0	1.05
22	PV	230	-	-	300	-60	96	0	1.05
23	PV	230	-	-	660	-125	310	0	1.05
24	PQ	230	-	-	-	-	-	0	-

Table A.4: Branches characteristics of the IEEE 24-bus test system.

Branch $i-j$ bus i	bus j	R [p.u.]	X [p.u.]	B_{sh} [p.u.]	a
1	2	0.0026	0.0139	0.4611	-
1	3	0.0546	0.2112	0.0572	-
1	5	0.0218	0.0845	0.0229	-
2	4	0.0328	0.1267	0.0343	-
2	6	0.0497	0.192	0.052	-
3	9	0.0308	0.119	0.0322	-
3	24	0.0023	0.0839	0	1.03
4	9	0.0268	0.1037	0.0281	-
5	10	0.0228	0.0883	0.0239	-
6	10	0.0139	0.0605	2.459	-
7	8	0.0159	0.0614	0.0166	-
8	9	0.0427	0.1651	0.0447	-
8	10	0.0427	0.1651	0.0447	-
9	11	0.0023	0.0839	0	1.03
9	12	0.0023	0.0839	0	1.03
10	11	0.0023	0.0839	0	1.02
10	12	0.0023	0.0839	0	1.02
11	13	0.0061	0.0476	0.0999	-
11	14	0.0054	0.0418	0.0879	-
12	13	0.0061	0.0476	0.0999	-
12	23	0.0124	0.0966	0.203	-
13	23	0.0111	0.0865	0.1818	-
14	16	0.005	0.0389	0.0818	-
15	16	0.0022	0.0173	0.0364	-
15	21	0.0126	0.098	0.206	-
15	24	0.0067	0.0519	0.1091	-
16	17	0.0033	0.0259	0.0545	-
16	19	0.003	0.0231	0.0485	-
17	18	0.0018	0.0144	0.0303	-
17	22	0.0135	0.1053	0.2212	-
18	21	0.0066	0.0518	0.1090	-
19	20	0.0102	0.0792	0.1666	-
20	23	0.0056	0.0432	0.0910	-
21	22	0.0087	0.0678	0.1424	-

A.2.2 Measurements Generation

Most of the test case studies of this thesis consider multiple measurements scenarios in order to derive a statistically-sound conclusion, as well as to generate database that is used for training and testing of neural networks. Each measurement scenario is generated from the solution of a converged power flow calculation and by adding Gaussian-distributed random errors to the obtained true values of active/reactive power and voltage magnitudes that are considered to be available. Distribution of the load level is determined by the data provided by [130], which follows the PDF shown in Figure A.4.

Hence, each measurements scenario generated for the 24-bus system involves the following:

- Load level determined from the load curve from figure A.4,
- Load on each bus is randomly modified from the base case within a range of $\pm 10\%$.
- On the converged power flow solutions zero-mean Gaussian errors were added, with standard deviation of 0.005 p.u. (for the system base power of 100 MVA) for both power and voltage magnitude measurements.
- If considering gross errors, they are generated uniformly with the magnitude of $20-200$ standard deviations of the Gaussian errors and added to a randomly chosen measurement.

Regarding the level of redundancy, i.e. measurements that are selected from the power flow results and assumed to be available for the state estimation process, various test samples were considered through the thesis.

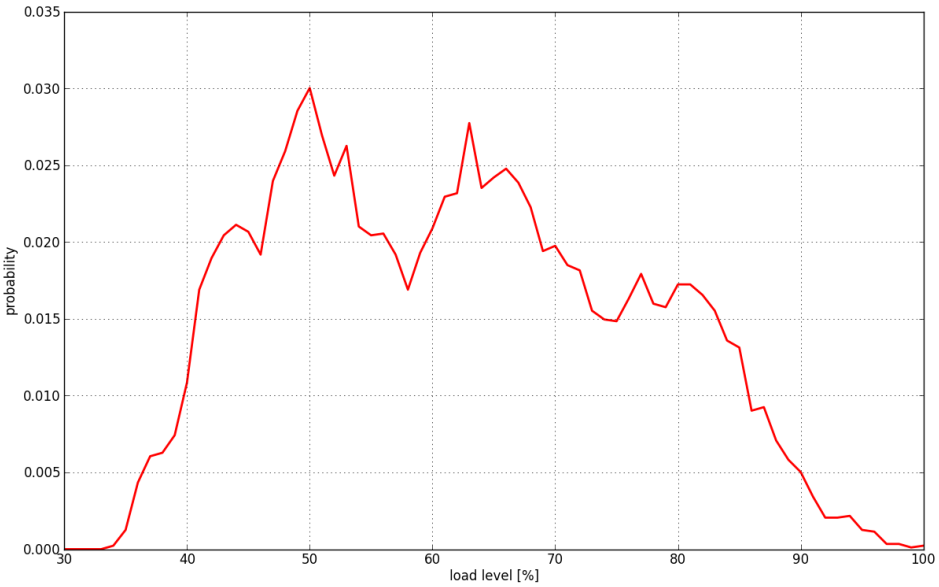


Figure A.4: PDF of load level used for the 24-bus system.

A.3 IEEE 118-bus Test System

Scalability of the proposed hybrid state estimation method, based on correntropy and auto-associative data screening, was tested on the IEEE 118-bus system, as the large-enough network. All the network details are available in [160]. Measurements generation was performed the same way as for the IEEE 24-bus system.

Appendix B

Information Theoretic Learning Criteria in PSSE

B.1 MSE and Correntropy - Differences and Similarities

The MSE criterion is defined for the two random variables X, Y and the error variable $E = Y - X$ as:

$$MSE(X,Y) = E[(X-Y)^2] = \iint_{x,y} (x-y)^2 f_{XY}(x,y)dxdy = \int_{e} e^2 f_E(e)de \tag{B.1}$$

where $f_{XY}(x,y)$ is joint PDF and $f_E(e)$ is the marginal distribution of errors. Therefore, in case of N observations of a variable, the Minimum Mean Square Error (MMSE) criterion, or simply least squares, the following term is minimized:

$$\frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2 = \frac{1}{N}\sum_{i=1}^{N}(e_i)^2 \tag{B.2}$$

The Correntropy function for the same variables X and Y is defined as:

$$\upsilon(X,Y) = E[G_\sigma(X-Y)] = \iint_{x,y} G_\sigma(x-y) f_{XY} dxdy = \int_{e} G_\sigma(e) f_E(e)de \tag{B.3}$$

The Information Theoretic Learning proposes estimation of Correntropy, directly from samples:

$$\hat{\upsilon}(X,Y) = \frac{1}{N}\sum_{i=1}^{N} G_\sigma(y_i - x_i) = \frac{1}{N}\sum_{i=1}^{N} G_\sigma(e_i) \tag{B.4}$$

Hereafter, MSE and Correntropy are compared as the two alternative similarity measures. The MSE quadratically increases by moving away from the $x = y$ line, so samples that are far away from the mean value of the error distribution will have an amplified impact on the function value. This explains why the MMSE estimator is optimal for the Gaussian distributed residuals, but also

why it is not suitable for the other error distributions, particularly if it is non-symmetric, has a nonzero mean or contains outliers.

Correntropy is also a similarity measure, but unlike the MSE which is definitely a global measure, it can be attributed as a local measure. Local properties are embedded by the Gaussian kernel bandwidth, which regulates the space around the $x = y$ line, from where the samples will contribute to the value of the similarity measure. Impact of the samples which are located outside of the local zone, defined by the kernel bandwidth σ, practically vanishes from the Correntropy function. These properties are particularly beneficial when Maximum Correntropy Criterion (MCC) is applied to errors with a non-Gaussian distribution.

Another property of the Correntropy function of key importance for the adoption as the novel state estimation criterion is the following: when the Parzen window (Gaussian kernel bandwidth) size is being increased, local similarity measure becomes global, i.e. Correntropy approaches quadratic function. In other words, when the Parzen window σ tends to infinity, the MCC estimator becomes equivalent to the MMSE (least squares) estimator. A proof for this property is provided hereafter.

If one looks at the equations (B.1) and (B.3) it is sufficient to prove that Gaussian function $G_\sigma(e)$ approaches quadratic function, when kernel bandwidth large enough [113]. Hence, assuming that the kernel size is very large, the errors of samples will be substantially smaller compared to the kernel size, thus allowing the second-order Taylor series expansion of the Gaussian function around zero to be a valid approximation:

$$G_\sigma(e) \approx G_\sigma(0) + G'_\sigma(0)e + G''_\sigma(0)e^2/2 \tag{B.5}$$

Also, due to the kernel function being symmetric and differentiable, its first-order derivative at zero is equal to zero, $G'_\sigma(0) = 0$, so it holds:

$$G_\sigma(e) \approx G_\sigma(0) + G''_\sigma(0)e^2/2 \tag{B.6}$$

Therefore, let $\sigma \to \sigma'$, $\sigma' \gg \max_i e_i$, i.e. let the Parzen window size of the Correntropy function be set to a value substantially greater than the maximum error. In that case, equation (B.6) holds, so MCC criterion can be written as the following:

$$\max_i \hat{\upsilon}(X,Y) = \max_i \frac{1}{N}\sum_{i=1}^{N} G_{\sigma'}(e_i) = \max_i \frac{1}{N}\sum_{i=1}^{N}(G_{\sigma'}(0) + G''_{\sigma'}(0)e_i^2/2) \tag{B.7}$$

The first term from equation (B.6) is $G_{\sigma'}(0) = 1/(\sigma\sqrt{2\pi})$, i.e. a positive constant. The second term contains second derivative of Gaussian kernel at 0, i.e. $G''_{\sigma'}(0)$ which is a negative constant since Gaussian is concave around zero. Therefore, it is straightforward to derive that maximization of the Correntropy function with $\sigma \to \sigma'$ is equivalent to minimization of the mean square error.

B.2 Mutual Information - a Criterion for Feature Selection

This thesis proposes a decentralized power system topology estimation based on auto-associative neural networks. In order the build a more efficient intelligent system, Chapter 6 proposes maximum mutual information as a criterion for selecting electric variables that are the most informative about a particular topology. This appendix brings a brief overview about using mutual information as a criterion for the feature selection problem.

As previously mentioned, the intention of a feature selector is to evaluate a set of candidate features and select a subset that is most relevant for a particular classification problem. The first work that suggests the adoption of MI as a criterion for selecting features for a general supervised neural network learning was published in [159]. Among many other feature selection methods, the reasons to select only the MI are multiple. There are various methods in the literature, mostly based on sensitivity analysis, linear transformations or principal component analysis. However, unlike the majority of well-accepted feature selection criteria, one of the key important properties of MI are capturing arbitrary relationship between variables and invariance under the invertible, linear space transformations. The main problem in that sense is calculation of MI in multivariate cases, due to excessive computational requirements. For example, it is well known that a group of k best features is not necessarily the best group of k features, since features are mostly not independent variables, and contain some information among themselves. In that regard, relevant research work has resulted in multiple variants of greedy algorithms that include also, in different ways, the mutual information between the measurements and minimize the redundant information for the classifier.

Battiti [159], as a pioneer in MI feature selection, suggested two terms that need to be used in order to substitute the calculation of multivariate mutual information $I(C,S)$ (between classes C and set of features S) for all the subsets $S \subset F$:

- $I(C,f_i)$ - mutual information between a particular feature f_i and the class variable C
- $I(f_i,f_s)$ - mutual information between individual candidate features f_i and individuals from the subset of selected features f_s.

With this simplification, computationally heavy calculation is substituted with a series of simply solvable terms, but with a certain loss in information measure accuracy. The proposed *Mutual Information Feature Selector* (MIFS) is in fact an incremental search greedy algorithm that selects features one by one. After the first feature is chosen, which corresponds to the maximum $I(C,f_i)$, further features are being selected by maximizing the following criterion:

$$I(C,f_i) - \beta \sum_{f_s \in S} I(f_i,f_s) \tag{B.8}$$

where parameter β regulates the relative importance of the MI between the candidate features and the group already-selected features S with respect to the MI with the output class. In other words,

this criterion intends to maximize information that feature contains about the classes and in the meantime minimize redundancy between the selected features.

This approach, initially proposed in [159] was later improved in articles [161]–[166]. Kwak and Choi in [162] proposed an enhancement of the MIFS method, called MIFS-U (*Mutual Information Feature Selector under Uniform Information Distribution*), which performs a better estimation of the mutual information between the input attributes and output classes. Additional representative improvements of these methods are Adaptive MIFS (AMIFS), proposed in [164], or the *min-redundancy max-relevance* (mRMR) criterion from [165]. All these feature selectors have in common the left-hand side term from the equation (B.8) which measures the individual relevance of the feature, and the right-hand side term which estimates the redundancy of the i^{th} feature with respect to the subset of previously selected features. Difference is in relationship between these two terms in order to minimize the approximation error from the multivariate MI. More specifically, the problem of the expression from eq. (B.8) is recognized in [166] due to incomparability of the two terms, and the user-defined parameter β which regulates their ratio. Accordingly, the solution is proposed using the fact that the MI between the two random variables is higher than zero and bounded from above by the minimum of their entropies:

$$0 \leq I(f_i, f_j) \leq \min\{H(f_i), H(f_j)\} \tag{B.9}$$

Accordingly, upper bound is proposed in [166] to normalize the MI between the candidate feature and the already selected features. The resulting criterion is called Normalized MIFS (NMIFS):

$$I(C, f_i) - \frac{1}{|S|} \sum_{f_s \in S} NI(f_i, f_s) \tag{B.10}$$

where $|S|$ is the size of the group of selected features and the NI stands for Normalized Mutual Information:

$$NI(f_i, f_s) = \frac{I(f_i, f_s)}{\min\{H(f_i), H(f_s)\}} \tag{B.11}$$

In this criterion (eq. (B.10)), the second term of redundancy penalization is adaptive and eliminates the necessity for the user-defined parameter β. The MI feature selection was most recently addressed in [167], [168]. Moreover, in [119], mutual information is introduced as a criterion for training of a pre-processing feature extractor, using the benefits of Information Theoretic Learning.

Appendix C

Auxiliary Methods

C.1 Line Search - the Armijo Rule

In section 3.2.5 the algorithm for the MCC state estimation is proposed. The core of the adopted solver is the Gauss-Newton method with an adaptive process of annealing the Parzen windows of the Correntropy function. Preliminary tests of Correntropy maximization have indicated eventual convergence problems, caused by inappropriate size of the iterative step.

Accordingly, in order to improve the search progress and eliminate convergence instability, a line-search optimization method which adapts the Newton's iterative step size is decided to be adopted. As it is reported in [169], large variety of methods are developed for optimizing the step size of the gradient search. Such methods are built on a trade-off between the fast progress in the objective function and time consumption needed to optimize the step size. Also, an important issue for performances of a line search algorithm is the terminating condition.

For the particular MCC state estimation with Gauss-Newton method, any additional computational demand is not welcome due to requirements of the real time feasibility. Therefore, a fast and not necessarily precise line search should be adopted. A very primitive step size update method called the *Backtracking Line Search* is used. More specifically, the initial step length is taken with $\alpha = 1$, and then it was being iteratively contracted until a termination criterion is satisfied. As a suitable criterion for that purpose the ***Armijo rule*** [170] is used, i.e. step size contraction is terminated if the following is satisfied:

$$J_{MCC}(x_k + \alpha\Delta x^k) > J_{MCC}(x_k) + c_1\alpha\left[\nabla J_{MCC}(x_k)\right]\Delta x^k \tag{C.1}$$

where c_1 is a constant defined on $\langle 0,1\rangle$ (it is usually set to a small value, in this thesis $c_1 = 10^{-4}$ is used) and α is the coefficient for the step size adaptation. Contraction of the step size α is being performed by the contraction factor ρ (in this thesis, $\rho = 2$ was used). Therefore, the algorithm used in this thesis for adaptation of the Newton's method step is the following:

Algorithm 4 Line Search

Choose $\alpha^0 > 0, \rho \in \langle 0,1 \rangle, c \in \langle 0,1 \rangle$;
Set $\alpha = \alpha^0$;
while $J_{MCC}(x_k + \alpha \Delta x^k) \leq J_{MCC}(x_k) + c_1 \alpha \left[\nabla J_{MCC}(x_k)\right] \Delta x^k$ **do**
    $\alpha^{k+1} = \rho \alpha^k$
Update the state vector by the Newton's step adapted by α^k:
$x^{k+1} = x^k + \alpha^k \Delta x^k$

C.2 Importance Sampling

In section 3.3.3, performances of the MCC estimator are tested on the 24-bus system. In order to derive a statistically-sound conclusion, an enormous number of GE test scenarios should be considered, what is barely feasible in some reasonable time. Therefore, the intention is to decrease required number of scenarios, without mitigating the statistical values. For that purpose, a technique from statistics is used called *importance sampling*.

Importance sampling relies on the distortion of the PDF (and Cumulative Density Function (CDF)), during which it increases the probability of important events at the cost of irrelevant ones. It also aims at reducing the variance without changing the expected value.

More specifically, let p be the PDF of failure, and a new distribution function p^* is called *importance distribution*. If the indicator function is $I = \{0 : \text{failure}, 1 : \text{success}\}$, importance sampling modifies it to the following:

$$I^* = \begin{cases} 0 \cdot \dfrac{1 - p_{fail}}{1 - p^*_{fail}}, & failure \\ 1 \cdot \dfrac{p_{fail}}{p^*_{fail}}, & success \end{cases} \tag{C.2}$$

Since the probability of estimation failure is not an *a priori* known value, it is assumed that failure is conditioned only with incidence and conformity of gross errors $p_{fail} = p_{incidence} \cdot p_{conformity}$. Also, probability that 2 GE are conforming is also very difficult to be quantified. Thus, distortion is performed to the probability of incidence between two measurements, instead of probability of failure. These assumptions are obviously not very precise and introduce certain losses in the scenarios number reduction. However, this should not be concerning since intentions of the performed analysis are more of comparative nature.

The probability of incidence is obtained by diving number of incident measurement couples with the total number of measurement couples:

$$p_{incidence} = \frac{N_{inc}}{\binom{m}{2}} = \frac{N_{inc}}{\dfrac{m!}{(m-2)! \cdot 2!}} = 10.65\% \tag{C.3}$$

where N_{inc} is total number of incident measurement couples. Incidence probability distortion is assumed to $p^*_{incidence} = 80\%$.

Now, the parameters and a stopping criterion of the Monte Carlo method need to be determined. Let $F(x)$ be a test function, which in this particular case measures estimator failure ($F(x) = I(x) = 0 : \text{failure}; 1 : \text{success}$). In the Monte Carlo method, the expectation of the test function $E(F)$ is estimated, based on a random sampling of N system states, or N scenarios:

$$\hat{E}(F) = \frac{1}{N}\sum_{i=1}^{N} F(x^i) \tag{C.4}$$

where x^i is i^{th} sampled vector from probability distribution $p(x)$ and $\hat{E}(F)$ is an estimate of the expectation $E(F)$. Variance of the sampling process $V(F)$ is also estimated as:

$$\hat{V}(F) = \frac{1}{N-1}\sum_{i=1}^{N}\left[F(x^i) - \bar{E}(F)\right]^2 \tag{C.5}$$

Obviously, the uncertainty in the estimate of $E(F)$ is inversely proportional to the sample size N and a stopping criterion for a Monte Carlo process is defined via the definition of a relative uncertainty, based on a variation coefficient β:

$$N = \frac{V(F)}{\beta \hat{E}(F)^2} \tag{C.6}$$

where $\beta^2 = \dfrac{V(\hat{E}(F))}{\hat{E}(F)^2}$. Therefore, the reduced number of samples N, for a given precision β, is directly conditioned with a variance $V(F)$. In a particular case study, when precision of $\beta = 5\%$ was assumed, it resulted in $N = 3760$ required samples.

Appendix D

The Relevant Published Work

The main part of the research work presented in this thesis resulted in publications published in scientific journals or international conferences. Some articles were submitted and are still under the revision process. Hereafter, a list of all relevant work can be found, according to the form as their are published, regardless of whether it has or is expected to be published.

Articles published in journals

- ◇ V. Miranda, J. Krstulovic, H. Keko, C. Moreira, and J. Pereira, "Reconstructing Missing Data in State Estimation With Autoencoders," IEEE Transactions on Power Systems, vol. 27, no. 2, pp. 604–611, 2012.
- ◇ J. Krstulovic, V. Miranda, A. Simões Costa, and J. Pereira, "Towards an Auto-Associative Topology State Estimator," IEEE Transactions on Power Systems, vol. 28, no. 3, pp. 3311–3318, 2013.

Papers submitted to journals

- ◇ C. Silva, J. Krstulovic, J. Hora, V. Palma, V. Miranda, and J. C. Principe "Extracting topology information from electric values: a comparison of models," submitted to IEEE Transactions on Power Systems
- ◇ V. Miranda, J. Krstulovic, and J. Pereira, "On the application of Correntropy to State Estimation in the presence of gross errors," in submission to IEEE Transactions on Power Systems
- ◇ V. Miranda, J. Krstulovic, and A. Simões Costa, "The statistical observability of topology in power networks," in submission to IEEE Transactions on Power Systems

Conference papers

- ◇ P. N. Pereira Barbeiro, J. Krstulovic, H. Teixeira, J. Pereira, F.J. Soares, J.P. Iria, "State Estimation in Distribution Smart Grids Using Autoencoders," PEOCO, Langkawi, Malaysia, 2014

- ◇ J. Krstulovic, V. Miranda, H. Keko, and J. Pereira, "Descoberta da topologia do sistema na ausência de sinais com redes neurais autoassociativas," IV Simpósio Brasileiro de Sistemas Elétricos, Goiânia, 2012.
- ◇ V. Miranda, J. Krstulovic, J. Hora, V. Palma, and J. C. Príncipe, "Breaker status uncovered by autoencoders under unsupervised maximum mutual information training," 17^{th} International Conference on Intelligent System Applications to Power Systems Tokyo, Japan, July 1-4, 2013.
- ◇ J. Krstulovic, and V. Miranda, "Selection of measurements in Topology Estimation with Mutual Information," submitted to IEEE International Energy Conference, Energycon, Dubrovnik, 2014.

Bibliography

[1] F. C. Schweppe and D. Rom, "Power System Static-State Estimation, Part II: Approximate Model", *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-89, no. 1, pp. 125–130, Jan. 1970.

[2] F. C. Schweppe and J. Wildes, "Power System Static-State Estimation, Part I: Exact Model", *IEEE Transactions on Power Apparatus and Systems*, no. 1, pp. 120–125, 1970.

[3] F. C. Schweppe, "Power system static-state estimation, Part III: Implementation", *IEEE Transactions on Power Apparatus and Systems*, no. 1, pp. 130–135, 1970.

[4] A. Abur and A. G. Expósito, *Power system state estimation: theory and implementation*, 1st. Marcel & Dekker Publishers, 2004, p. 327.

[5] R. van Amerongen, "On convergence analysis and convergence enhancement of power system least-squares state estimators", *IEEE Transactions on Power Systems*, vol. 10, no. 4, pp. 2038–2044, 1995.

[6] K. A. Clements, G. R. Krumpholz, and P. W. Davis, "Power system state estimation residual analysis: an algorithm using network topology", *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-100, no. 4, pp. 1779–1787, 1981.

[7] G. Korres and G. Contaxis, "A reduced model for bad data processing in state estimation", *IEEE Transactions on Power Systems*, vol. 6, no. 2, pp. 550–557, 1991.

[8] H. M. Merrill and F. C. Schweppe, "Bad data suppression in power system static state estimation", *IEEE Transactions on Power Apparatus and Systems*, no. 6, pp. 2718–2725, 1971.

[9] E. Handschin, F. C. Schweppe, J. Kohlas, and A. Fiechter, "Bad Data Analysis for Power System State Estimation", *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-94, no. 2, pp. 329–337, 1975.

[10] F. Zhuang and R. Balasubramanian, "Bad Data Suppression in Power System State Estimation with a Variable Quadratic-Constant Criterion", *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-104, no. 4, pp. 857–863, Jul. 1985.

[11] A. Simões Costa, J. Rolim, and P. Aitchison, "Iterative bad-data suppression applied to state estimators based on the augmented matrix method", *Electric Power Systems Research*, vol. 20, no. 3, pp. 205–213, 1991.

[12] R. Baldick, K. A. Clements, Z. Pinjo-Dzigal, and P. W. Davis, "Implementing Non—Quadratic Objective Functions for State Estimation and Bad Data Rejection", *IEEE Transactions on Power Systems*, vol. 12, no. 1, 1997.

[13] Y. Ejima, H. Kondo, and S. Iwamoto, "New Bad Data Rejection Algorithm using Nonquadratic Objective Function for State Estimation", *Power Engineering Society General Meeting, IEEE*, vol. 2, no. 2, pp. 1–8, 2007.

[14] M. R. Irving, R. C. Owen, and J. H. Sterling, "Power System State Estimation Using Linear Programming", *Proceedings of IEEE*, vol. 125, pp. 879–885, 1978.

[15] W. W. Kotiuga and M. Vidyasagar, "Bad Data rejection properties of weighted least absolute value techniques applied to static state estimation", *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-101, no. 4, pp. 844–853, 1982.

[16] D. Falcão and S. de Assis, "Linear Programming State Estimation: Error Analysis and Gross Error Identification", *IEEE Transactions on Power Systems*, vol. 3, no. 3, 1988.

[17] L. Mili, T. V. Cutsem, and M. Ribbens-Pavella, "Hypothesis testing identification: A new method for bad data analysis in power system state estimation", *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-103, no. 11, pp. 3239–3252, 1984.

[18] L. Mili, T. V. Cutsem, and M. Ribbens-Pavella, "Bad data identification methods in power system state estimation - a comparative study", *Power*, vol. PAS-104, no. 11, pp. 3037–3049, 1985.

[19] M. K. Çelik and A. Abur, "A robust WLAV state estimator using transformations", *IEEE Transactions on Power Systems*, vol. 7, no. 1, pp. 106–113, 1992.

[20] M. K. Çelik and A. Abur, "Use of scaling in WLAV estimation of power system states", *IEEE Transactions on Power Systems*, vol. 7, no. 2, pp. 684–692, 1992.

[21] L. Mili, M. G. Cheniae, and S. Vichare, "Robust State estimation based on projection statistics", *IEEE Transactions on Power Systems*, vol. 11, no. 2, 1996.

[22] P. J. Huber, *Robust Statistics*, ser. Wiley Series in Probability and Statistics. New York, USA: John Wiley & Sons, Inc., Feb. 1981.

[23] P. Rousseeuw and A. Leroy, *Robust regression and outlier detection*. John Wiley & Sons, Inc., 1987.

[24] A. Monticelli, F. F. Wu, and M. Yen, "Multiple bad data identification for state estimation by combinatorial optimization", *IEEE Transactions on Power Delivery*, vol. PWRD-1, no. 3, pp. 361–369, 1986.

[25] P. J. Rousseeuw, "Least Median of Squares Regression", *Journal of the American Statistical Association*, vol. 79, no. 388, p. 871, Dec. 1984.

[26] L. Mili, V. Phaniraj, and P. Rousseeuw, "Least median of squares estimation in power systems", *IEEE Transactions on Power Systems*, vol. 6, no. 2, pp. 511–523, 1991.

[27] L. Mili, M. Cheniae, and P. Rousseeuw, "Robust state estimation of electric power systems", *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 41, no. 5, pp. 349–358, May 1994.

[28] S. Gastoni, G. Granelli, and M. Montagna, "Robust State-Estimation Procedure Based on the Maximum Agreement Between Measurements", *IEEE Transactions on Power Systems*, vol. 19, no. 4, pp. 2038–2043, Nov. 2004.

[29] E. N. Asada, A. V. Garcia, and R. Romero, "Identifying multiple interacting bad data in power system state estimation", *IEEE Power Engineering Society General Meeting, 2005*, no. 5, pp. 3013–3019, 2005.

[30] J. Khwanram and P. Damrongkulkamjorn, "Multiple bad data identification in power system state estimation using particle swarm optimization", *2009 6th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, vol. 1, no. 3, pp. 2–5, May 2009.

[31] A. Al-Othman and M. Irving, "Robust state estimator based on maximum constraints satisfaction of uncertain measurements", *Measurement*, vol. 40, no. 3, pp. 347–359, Apr. 2007.

[32] M. Irving, "Robust state estimation using mixed integer programming", *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 1519–1520, 2008.

[33] M. Irving, "Robust Algorithm for Generalized State Estimation", *IEEE Transactions on Power Systems*, vol. 24, no. 4, pp. 1886–1887, Nov. 2009.

[34] Y. Weng, R. Negi, Q. Liu, and M. Ilic, "Robust state-estimation procedure using a Least Trimmed Squares pre-processor", *Innovative Smart Grid Technologies, 2011 IEEE PES*, pp. 1–6, 2011.

[35] F. Hampel, E. Ronchetti, P. Rousseeuw, and W. Stahel, *Robust Statistics: The Approach Based on Influence Functions*. Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., 1986.

[36] D. M. Falcão and A. Brameller, "Power system tracking state estimation and bad data processing", *IEEE Transactions on Power Apparatus and Systems*, no. 2, pp. 325–333, 1982.

[37] L. Mili and T. V. Cutsem, "Implementation of the Hypothesis Testing Identification in Power System State Estimation", *IEEE Transactions on Power Systems*, vol. 3, no. 3, pp. 887–893, 1988.

[38] N. G. Bretas, J. B. A. London Jr., L. F. C. Alberto, and R. A. S. Benedito, "Geometrical Approach on Masked Gross Errors for Power Systems State Estimation", *Matrix*, pp. 1–7, 2009.

[39] N. Bretas and J. B. A. London Jr., "Recovering of masked errors in power systems state estimation and measurement gross error detection and identification proposition", *Power and Energy Society General Meeting, 2010 IEEE*, pp. 1–6, 2010.

[40] N. Bretas and S. Pierreti, "The innovation concept in bad data analysis using the composed measurements errors for power system state estimation", in *Power and Energy Society General Meeting, 2010 IEEE*, IEEE, 2010, pp. 1–6.

[41] N. G. Bretas, A. S. Bretas, and S. A. R. Piereti, "Masked errors in power systems state estimation and measurement gross errors detection and identification", *PowerTech, 2011 IEEE Trondheim*, pp. 1–5, 2011.

[42] G. He, S. Dong, J. Qi, and Y. Wang, "Robust State Estimator Based on Maximum Normal Measurement Rate", *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2058–2065, 2011.

[43] E. Caro, A. J. Conejo, and R. Mínguez, "Power system state estimation considering measurement dependencies", *IEEE Transactions on Power Systems*, vol. 24, no. 4, pp. 1875–1885, 2009.

[44] E. Caro, J. Morales, A. J. Conejo, and R. Mínguez, "Calculation of measurement correlations using point estimate", *IEEE Transactions on Power Delivery*, vol. 25, no. 4, pp. 2095–2103, 2010.

[45] E. Caro, A. J. Conejo, R. Mínguez, M. Zima, and G. Andersson, "Multiple bad data identification considering measurement dependencies", *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 1953–1961, 2011.

[46] J. Chen and A. Abur, "Improved bad data processing via strategic placement of PMUs", *Power Engineering Society General Meeting,*, pp. 1–5, 2005.

[47] J. Chen and A. Abur, "Placement of PMUs to enable bad data detection in state estimation", *IEEE Transactions on Power Systems*, vol. 21, no. 4, pp. 1608–1615, 2006.

[48] J. Zhu and A. Abur, "Bad data identification when using phasor measurements", *Power Tech, 2007 IEEE Lausanne*, pp. 1676–1681, 2007.

[49] J. Zhu and A. Abur, "Identification of network parameter errors using phasor measurements", *2009 IEEE Power & Energy Society General Meeting*, pp. 1–5, Jul. 2009.

[50] A. Abur, "Impact of phasor measurements on state estimation", *Electrical and Electronics Engineering, 2009. ELECO ...*, pp. 3–7, 2009.

[51] G. N. Korres and N. M. Manousakis, "State estimation and bad data processing for systems including PMU and SCADA measurements", *Electric Power Systems Research*, vol. 81, no. 7, pp. 1514–1524, Jul. 2011.

[52] H. Xue, Q. Jia, N. Wang, Z. Bo, H. Wang, and H. Ma, "A Dynamic State Estimation method with PMU and SCADA measurement for power systems", in *Power Engineering Conference, 2007. IPEC 2007. International*, vol. 5, 2007, pp. 848–853.

[53] A. Jain and N. R. Shivakumar, "Phasor measurements in dynamic state estimation of power systems", in *TENCON 2008 - 2008 IEEE Region 10 Conference*, 2008, pp. 1–6.

[54] E. Ghahremani and I. Kamwa, "Dynamic State Estimation in Power System by Applying the Extended Kalman Filter With Unknown Inputs to Phasor Measurements", *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2556–2566, 2011.

[55] A. Gomez-Exposito and A. de la Villa Jaen, "Two-Level State Estimation With Local Measurement Pre-Processing", *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 676–684, 2009.

[56] C. Gomez-Quiles, A. de la Villa Jaen, and A. Gomez-Exposito, "A Factorized Approach to WLS State Estimation", *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1724–1732, 2011.

[57] T. Yang, H. Sun, and A. Bose, "Transition to a Two-Level Linear State Estimator — Part I : Architecture", *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 46–53, 2011.

[58] T. Yang, H. Sun, and A. Bose, "Transition to a Two-Level Linear State Estimator — Part II : Algorithm", *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 54–62, 2011.

[59] A. Gomez-Exposito, A. Abur, A. de la Villa Jaen, and C. Gomez-Quiles, "A Multilevel State Estimation Paradigm for Smart Grids", *Proceedings of the IEEE*, vol. 99, no. 6, pp. 952–976, 2011.

[60] C. Gomez-Quiles, A. Gomez-Exposito, and A. de la Villa Jaen, "State Estimation for Smart Distribution Substations", *IEEE Transactions on Smart Grid*, vol. 3, no. 2, pp. 986–995, 2012.

[61] A. Bose, "Smart transmission grid applications and their supporting infrastructure", *IEEE Transactions on Smart Grid*, vol. 1, no. 1, pp. 11–19, Jun. 2010.

[62] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious Data Attacks on Smart Grid State Estimation: Attack Strategies and Countermeasures", *2010 First IEEE International Conference on Smart Grid Communications*, pp. 220–225, Oct. 2010.

[63] Y. Hu, A. Kuh, A. Kavcic, and D. Nakafuji, "Real-time state estimation on micro-grids", *The 2011 International Joint Conference on Neural Networks*, pp. 1378–1385, Jul. 2011.

[64] A. Debs and R. E. Larson, "A dynamic estimator for tracking the state of a power system", *IEEE Transactions on Power Apparatus and Systems*, no. 7, pp. 1670–1678, 1970.

[65] K. Srinivasan and Y. Robichaud, "A dynamic estimator for complex bus voltage determination", *IEEE, Power Apparatus and Systems*, pp. 1581–1588, 1974.

[66] R. D. Masiello and F. C. Schweppe, "A tracking static state estimator", *IEEE Transactions on Power Apparatus and Systems*, no. x, pp. 1025–1033, 1970.

[67] K. Nishiya, J. Hasegawa, B. Eng, M. Eng, P. T. Koike, and D. Eng, "Dynamic state estimation including anomaly detection and identification for power systems", *Proceedings of IEEE*, vol. 129, no. 5, 1982.

[68] A. M. L. Silva, M. B. Coutto Filho, and J. Cantera, "An efficient dynamic state estimation algorithm including bad data processing", *IEEE Transactions on Power Systems*, no. 4, pp. 1050–1058, 1987.

[69] A. Alves da Silva, V. Quintana, and G. Pang, "Solving data acquisition and processing problems in power systems using a pattern analysis approach", *IEEE Proceedings-Generation, Transmission and Distribution*, vol. 138, no. 4, pp. 365–376, 1991.

[70] A. Alves da Silva and V. Quintana, "Pattern analysis in power system state estimation", *International Journal of Electrical Power & Energy Systems*, vol. 17, no. 1, pp. 51–60, 1995.

[71] J. C. S. Souza, A. L. da Silva, and A. A. da Silva, "Data debugging for real-time power system monitoring based on pattern analysis", *IEEE Transactions on Power Systems*, vol. 11, no. 3, 1996.

[72] J. C. S. Souza, A. Leite da Silva, and A. Alves da Silva, "Data visualisation and identification of anomalies in power system state estimation using artificial neural networks", *IEEE Proceedings-Generation, Transmission and Distribution*, vol. 144, pp. 445–455, 1997.

[73] J. Souza, A. L. da Silva, and A. A. da Silva, "Online topology determination and bad data suppression in power system operation using artificial neural networks", *IEEE Transactions on Power Systems*, vol. 13, no. 3, pp. 796–803, 1998.

[74] A. K. Sinha and J. K. Mandal, "Dynamic state estimator using ANN based bus load prediction", *IEEE Transactions on Power Systems*, vol. 14, no. 4, pp. 1219–1225, 1999.

[75] A. Sinha and J. K. Mandal, "Hierarchical dynamic state estimator using ANN-based dynamic load prediction", *IEEE Proceedings-Generation, Transmission and Distribution*, vol. 146, no. 6, p. 541, 1999.

[76] K. Shih and S. Huang, "Application of a robust algorithm for dynamic state estimation of a power system", *IEEE Transactions on Power Systems*, vol. 17, no. 1, pp. 141–147, 2002.

[77] J. Lin, S. Huang, and K. Shih, "Application of sliding surface-enhanced fuzzy control for dynamic state estimation of a power system", *IEEE Transactions on Power Systems*, vol. 18, no. 2, pp. 570–577, 2003.

[78] S. Huang and J.-m. Lin, "Enhancement of anomalous data mining in power system predicting-aided state estimation", *IEEE Transactions on Power Systems*, vol. 19, no. 1, pp. 610–619, 2004.

[79] M. B. Do Coutto Filho, J. C. S. Souza, R. S. G. Matos, and M. T. Schilling, "Revealing gross errors in critical measurements and sets via forecasting-aided state estimators", *Electric Power Systems Research*, vol. 57, no. 1, pp. 25–32, Jan. 2001.

[80] B. M. do Coutto Filho and J. C. Stacchini De Souza, "Forecasting-aided state estimation—Part I: Panorama", *IEEE Transactions on Power Systems*, vol. 24, no. 4, pp. 1667–1677, 2009.

[81] B. M. do Coutto Filho and J. C. Stacchini De Souza, "Forecasting-Aided State Estimation — Part II: Implementation", vol. 24, no. 4, pp. 1678–1685, 2009.

[82] R. Lugtu, D. Hackett, K. Liu, and D. Might, "Power system state estimation: Detection of topological errors", *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-99, no. 6, pp. 2406–2412, 1980.

[83] K. Clements and P. W. Dawis, "Detection and identification of topology errors in electric power systems", *IEEE Transactions on Power Systems*, vol. 3, no. 4, pp. 1748–1753, 1988.

[84] F. F. Wu and W. E. Liu, "Detection of topology errors by state estimation", *IEEE Transactions on Power Systems*, vol. 4, no. 1, pp. 176–183, 1989.

[85] A. Simões Costa and J. A. Leão, "Identification of topology errors in power system state estimation", *IEEE Transactions on Power Systems*, vol. 8, no. 4, pp. 1531–1538, 1993.

[86] M. Prais and A. Bose, "A topology processor that tracks network modification over time", *IEEE Transactions on Power Systems*, vol. 3, no. 3, pp. 992–998, 1988.

[87] M. Irving and J. H. Steriling, "Substation data validation", *IEEE Proceedings-Generation, Transmission and Distribution*, vol. 129, no. 3, pp. 119–122, 1982.

[88] N. Singh and H. Glavitsch, "Detection and identification of topological errors in online power system analysis", *IEEE Transactions on Power Systems*, no. 1, pp. 324–331, 1991.

[89] N. Singh and F. Oesch, "Practical Experience with Rule-based On-line Topology Error Detection", *IEEE Transactions on Power Systems*, vol. 9, no. 2, 1994.

[90] A. Monticelli, "Modeling zero impedance branches in power system state estimation", *IEEE Transactions on Power Systems*, vol. 6, no. 4, 1991.

[91] A. Monticelli, "Modeling circuit breakers in weighted least squares state estimation", *IEEE Transactions on Power Systems*, vol. 8, no. 3, pp. 1143–1149, 1993.

[92] O. Alsac, N. Vempati, B. Stott, and a. Monticelli, "Generalized state estimation", *IEEE Transactions on Power Systems*, vol. 13, no. 3, pp. 1069–1075, 1998.

[93] A. Abur, H. Kim, and M. Celik, "Identifying the unknown circuit breaker statuses in power networks", *IEEE Transactions on Power Systems*, vol. 10, no. 4, pp. 2029–2037, 1995.

[94] H. Singh and F. L. Alvarado, "Network topology determination using Least Absolute Value State Estimation", *IEEE Transactions on Power Systems*, vol. 10, no. 3, 1995.

[95] L. Mili, G. Steeno, F. Dobraca, and D. French, "A robust estimation method for topology error identification", *IEEE Transactions on Power Systems*, vol. 14, no. 4, 1999.

[96] K. A. Clements and A. Simoes Costa, "Topology error identification using normalized Lagrange multipliers", *IEEE Transactions on Power Systems*, vol. 13, no. 2, pp. 347–353, 1998.

[97] E. M. Lourenço, A. Simões Costa, and K. A. Clements, "Bayesian-Based Hypothesis Testing for Topology Error Identification in Generalized State Estimation", *IEEE Transactions on Power Systems*, vol. 19, no. 2, pp. 1206–1215, 2004.

[98] E. M. Lourenço, A. Simoes Costa, K. A. Clements, and R. A. Cernev, "A topology error identification method directly based on collinearity tests", *IEEE Transactions on Power Systems*, vol. 21, no. 4, pp. 1920–1929, 2006.

[99] J. Pereira, "A State Estimation Approach for Distribution Networks Considering Uncertainties and Switching", PhD Thesis, Faculty of Engineering of the University of Porto, Portugal, 2001.

[100] J. Pereira, V. Miranda, and J. Saraiva, "Fuzzy Control of State Estimation Robustness", *Proceedings of the 14th PSCC, Seville, Spain*, 2002.

[101] J. Pereira, J. Saraiva, V. Miranda, A. Costa Simões, E. M. Lourenço, and K. A. Clements, "Comparison of approaches to identify topology errors in the scope of state estimation studies", in *Power Tech Proceedings, 2001 IEEE Porto*, vol. 3, IEEE, 2001, 6–pp.

[102] E. Caro, A. J. Conejo, and A. Abur, "Breaker status identification", *IEEE Transactions on Power Systems*, vol. 25, no. 2, pp. 694–702, 2010.

[103] A. Alves da Silva, V. Quintana, and G. Pang, "A pattern analysis approach for topology determination, bad data correction and missing measurement estimation in power systems", *Power Symposium, 1990. Proceedings of the Twenty-Second Annual North American*, pp. 363 –372, 1990.

[104] A. P. A. Silva, V. H. Quintana, and G. K. H. Pang, "Neural networks for topology determination of power systems", 1991.

[105] D. Vinod Kumar, S. Srivastava, S. Shah, and S. Mathur, "Topology processing and static state estimation using artificial neural networks", in *IEEE Proceedings-Generation, Transmission and Distribution*, vol. 143, IET, 1996, pp. 99–105.

[106] C. Lu, J. Teng, and B. Chang, "Power system network topology error detection", *IEEE Proceedings-Generation, Transmission and Distribution*, vol. 141, no. 6, 1994.

[107] T. Tian and M. Zhu, "An artificial neural network-based expert system for network topological error identification", *Neural Networks, 1995. Proceedings*, vol. 2, pp. 882–886, 1995.

[108] R. Lukomski, "Identification of topology errors with use of unbalance indices and neural networks", *PowerTech, 2009 IEEE Bucharest*, pp. 1–8, Jun. 2009.

[109] D. Singh and J. Pandey, "Topology identification, bad data processing, and state estimation using fuzzy pattern matching", *IEEE Transactions on Power Systems*, 2005.

[110] A. Monticelli, *State estimation in electric power systems: A Generalized Approach*, 1st, M. A. Pai, Ed. Kluwer Academic Publishers, 1999.

[111] V. Miranda, A. Santos, and J. Pereira, "State estimation based on correntropy: a proof of concept", *IEEE Transactions on Power Systems*, vol. 24, no. 4, pp. 1888–1889, Nov. 2009.

[112] C. Shannon, "A mathematical theory of communication", *The Bell System Technical Journal*, vol. 27, pp. 379–423, 623656–, 1948.

[113] J. C. Principe, *Information theoretic learning: Renyi's Entropy and Kernel Perspectives*, 1st. Springer, 2010, vol. 1.

[114] D. Xu, J. Principe, J. Fisher, and H.-C. Wu, "A novel measure for independent component analysis (ICA)", *Proceedings of ICASSP'98*, vol. 2, pp. 1161–1164, 1998.

[115] J. C. Principe, D. Xu, and J. W. Fisher, "Information Theoretic Learning", in *Unsupervised adaptive filtering*, New York: Wiley, 2000, pp. 265–319.

[116] A. Renyi, "Some Fundamental Questions of Information Theory", *Selected Papers of Alfred Renyi, Academia Kiado, Budapest*, vol. 2, pp. 526–552, 1976.

[117] E. Parzen, "On estimation of a probability density function and mode", *The annals of mathematical statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.

[118] D. Erdogmus and J. C. Principe, "An error-entropy minimization algorithm for supervised training of nonlinear adaptive systems", *IEEE Transactions on Signal Processing*, vol. 50, no. 7, pp. 1780–1786, 2002.

[119] K. Hild, D. Erdogmus, K. Torkkola, and J. C. Principe, "Feature extraction using information-theoretic learning", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 9, pp. 1385–1392, 2006.

[120] I. Santamaría, P. P. Pokharel, and J. C. Principe, "Generalized correlation function: definition, properties, and application to blind equalization", *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 2187–2197, 2006.

[121] W. Liu, P. Pokharel, and J. Principe, "Correntropy: properties and applications in non-Gaussian signal processing", *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5286–5298, 2007.

[122] W. Liu, P. Pokharel, and J. Principe, "Correntropy: A localized similarity measure", in *Neural Networks, 2006. IJCNN'06. International Joint Conference on*, IEEE, 2006, pp. 4919–4924.

[123] V. Miranda and N. Fonseca, "EPSO – Best-of-two-worlds metaheuristic applied to power system problems", in *Proceedings of World Congress on Computational Intelligence*, Honolulu (Hawaii), 2002.

[124] V. Miranda and N. Fonseca, "EPSO-evolutionary particle swarm optimization, a new algorithm with applications in power systems", *Transmission and Distribution Conference and Exhibition 2002: Asia Pacific. IEEE/PES*, vol. 2, pp. 745–750, 2002.

[125] V. Miranda, H. Keko, and A. Jaramillo, "EPSO: Evolutionary Particle Swarms", in *Advances in Evolutionary Computing for System Design*, vol. 167, Springer, series: Studies In Computational Intelligence, 2007, ch. 6, pp. 139–168.

[126] R. Jenssen, D. Erdogmus, K. E. Hild, J. C. Principe, and T. r. Eltoft, "Information cut for clustering using a gradient descent approach", *Pattern Recognition*, vol. 40, no. 3, pp. 796–806, Mar. 2007.

[127] A. Singh and J. C. Principe, "Kernel width adaptation in information theoretic cost functions", in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, IEEE, 2010, pp. 2062–2065.

[128] J. J. Grainger and W. D. Stevenson, *Power System Analysis*. McGraw-Hill, Inc., 1994.

[129] E. Caro and A. J. Conejo, "State estimation via mathematical programming: a comparison of different estimation algorithms", *IET Generation, Transmission & Distribution*, vol. 6, no. 6, p. 545, 2012.

[130] *IEEE PES Task Force, IEEE reliability test system*, 1979. [Online]. Available: `http://pscal.ece.gatech.edu/testsys/index.html`.

[131] V. Miranda, "Principles of Monte Carlo simulation", Tech. Rep., 2010.

[132] W. Liu, P. Pokharel, and J. Principe, "Error entropy, correntropy and m-estimation", *Machine Learning for Signal Processing, 2006. Proceedings of the 2006 16th IEEE Signal Processing Society Workshop on*, vol. 2, no. 3, pp. 1–6, 2006.

[133] K. Lo, P. Ong, R. McColl, A. Moffatt, and J. Sulley, "Development of a Static State Estimator Part I: Estimation and Bad Data Suppression", *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-102, no. 8, pp. 2486–2491, 1983.

[134] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks.", *Science (New York, N.Y.)*, vol. 313, no. 5786, pp. 504–7, Jul. 2006.

[135] G. Cottrell, P. Munro, and D. Zipser, "Learning internal representations from gray-scale images: An example of extensional programming", *Proc. 9th Annu. Conf. Cognitive Science Society, Seattle, WA*, 1987.

[136] A. Basso, "Autoassociative neural networks for image compression: a massively parallel implementation", *Neural Networks for Signal Processing [1992] II., Proceedings of the 1992 IEEE-SP Workshop*, pp. 373–381, 1992.

[137] M. Fleming and G. Cottrell, "Categorization of faces using unsupervised feature extraction", *Proc. IJCNN—Int. Joint Conf. Neural Networks, San Diego, CA, Jun. 17–21*, vol. 2, pp. 65–70, 1990.

[138] B. Golomb and T. Sejnowski, "Sex recognition from faces using neural networks", *Application of Neural Networks, Murray, A.(Editor),*, pp. 71–92, 1995.

[139] S. Narayanan, R. Marks, J. Vian, J. Choi, M. El-Sharkawi, and B. Thompson, "Set constraint discovery: missing sensor data restoration using autoassociative regression machines", in *Proceedings of the 2002 International Joint Conference on Neural Networks*, Honolulu (Hawaii), 2002, pp. 2872–2877.

[140] B. Thompson, R. Marks, and M. El-Sharkawi, "On the contractive nature of autoencoders: Application to missing sensor restoration", *Proceedings of the International Joint Conference on Neural Networks*, vol. 4, no. 1, pp. 3011–3016, 2003.

[141] I. T. Jolliffe, *Principal Component Analysis*, 2nd. Springer Series in Statistics, Springer, 2002.

[142] M. a. Kramer, "Nonlinear principal component analysis using autoassociative neural networks", *AIChE Journal*, vol. 37, no. 2, pp. 233–243, Feb. 1991.

[143] N. Japkowicz, S. Hanson, H. Mark, and M. A. Gluck, "Nonlinear autoassociation is not equivalent to PCA", *Neural Computation*, vol. 12, no. 3, pp. 531–545, 2000.

[144] M. Abdella and T. Marwala, "The use of genetic algorithms and neural networks to approximate missing data in database", *IEEE 3rd International Conference on Computational Cybernetics, 2005. ICCC 2005.*, pp. 207–212, 2005.

[145] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders", *Proceedings of the 25th international conference on Machine learning - ICML '08*, pp. 1096–1103, 2008.

[146] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion", *The Journal of Machine Learning Research*, vol. 11, pp. 3371–3408, 2010.

[147] M. Chen, Z. Xu, K. Q. Weinberger, and F. Sha, "Marginalized denoising autoencoders for domain adaptation", *Proceedings of the 29th International Conference on Machine Learning, Edinburgh, Scotland, UK*, 2012.

[148] W. Qiao, Z. Gao, R. Harley, and G. K. Venayagamoorthy, "Robust neuro-identification of nonlinear plants in electric power systems with missing sensor measurements", *Engineering Applications of Artificial Intelligence*, vol. 21, no. 4, pp. 604–618, 2008.

[149] S. Mohagheghi, G. Venayagamoorthy, and R. Harley, "Optimal wide area controller and state predictor for a power system", *IEEE Transactions on Power Systems*, vol. 22, no. 2, pp. 693–705, 2007.

[150] V. Miranda, A. R. Garcez Castro, and S. Lima, "Diagnosing faults in power transformers with autoassociative neural networks and mean shift", *IEEE Transactions on Power Delivery*, vol. 27, no. 3, pp. 1350–1357, 2012.

[151] H. Salehfar and R. Zhao, "A neural network preestimation filter for bad-data detection and identification in power system state estimation", *Electric power systems research*, vol. 34, pp. 127–134, 1995.

[152] N. Abbasy and W. El-Hassawy, "Power system state estimation: ANN application to bad data detection and identification", *AFRICON, 1996., IEEE AFRICON 4th*, vol. 2, pp. 611 –615, 1996.

[153] S. Teeuwsen and I. Erlich, "Neural network based multi-dimensional feature forecasting for bad data detection and feature restoration in power systems", *Power Engineering Society General Meeting, 2006. IEEE*, 2006.

[154] V. Miranda, J. Krstulovic, H. Keko, C. Moreira, and J. Pereira, "Reconstructing Missing Data in State Estimation With Autoencoders", *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 604–611, May 2012.

[155] A. Simões Costa and M. Tardio Arze, "Critical pseudo-measurement selection for unreduced external system modeling", *International Journal of Electrical Power & Energy Systems*, vol. 18, no. 2, pp. 73–80, Feb. 1996.

[156] A. Simões Costa and M. Carneiro dos Santos, "Real-Time Monitoring of Distributed Generation Based on State Estimation and Hypothesis Testing", *2007 IEEE Lausanne Power Tech*, pp. 538–543, Jul. 2007.

[157] E. Manitsas, R. Singh, B. Pal, and G. Strbac, "Modeling of pseudo-measurements for distribution system state estimation", in *SmartGrids for Distribution, 2008. IET-CIRED. CIRED Seminar*, 2008, pp. 1–4.

[158] J. Krstulovic, V. Miranda, A. Simões Costa, and J. Pereira, "Towards an Auto-Associative Topology State Estimator", *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 3311–3318, 2013.

[159] R. Battiti, "Using mutual information for selecting features in supervised neural net learning.", *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, vol. 5, no. 4, pp. 537–50, Jan. 1994.

[160] *Power Systems Test Case Archive*, 2013. [Online]. Available: `http://www.ee.washington.edu/research/pstca/pf118/pg_tca118bus.htm`.

[161] N. Kwak and C.-h. Choi, "Improved mutual information feature selector for neural networks in supervised learning", *Neural Networks, 1999. IJCNN '99. International Joint Conference on*, vol. 2, pp. 1313 –1318, 1999.

[162] N. Kwak and C.-H. Choi, "Input feature selection for classification problems", *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, vol. 13, no. 1, pp. 143–59, Jan. 2002.

[163] N. Kwak and C. Choi, "Input feature selection by mutual information based on Parzen window", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 12, pp. 1667–1671, 2002.

[164] M. Tesmer and P. Estévez, "Amifs: adaptive feature selection by using mutual information", *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, vol. 1, pp. 1–6, 2004.

[165] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[166] P. a. Estévez, M. Tesmer, C. a. Perez, and J. M. Zurada, "Normalized mutual information feature selection.", *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, vol. 20, no. 2, pp. 189–201, Feb. 2009.

[167] L. T. Vinh, N. D. Thang, and Y.-K. Lee, "An Improved Maximum Relevance and Minimum Redundancy Feature Selection Algorithm Based on Normalized Mutual Information", *2010 10th IEEE/IPSJ International Symposium on Applications and the Internet*, no. 3, pp. 395–398, Jul. 2010.

[168] L. T. Vinh, S. Lee, Y.-T. Park, and B. J. D'Auriol, "A novel feature selection method based on normalized mutual information", *Applied Intelligence*, vol. 37, no. 1, pp. 100–120, Aug. 2012.

[169] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd. Springer Series in Operations Research and Financial Engineering, 2006.

[170] L. Armijo, "Minimization of functions having Lipschitz continuous first partial derivatives.", *Pacific Journal of mathematics*, vol. 16, no. 1, pp. 1–3, 1966.

BIBLIOGRAPHY