

Igor Ferreira Visconti

A Load Modeling Methodology for Steady State and Dynamic Simulations

Tese de Doutorado

Thesis presented to the Programa de Pós-graduação
em Engenharia Elétrica of PUC-Rio in partial
fulfillment of the requirements for the degree of Doutor
em Engenharia Elétrica.

Advisor: Prof. Delberis Araujo Lima

Rio de Janeiro

July 2019

Igor Ferreira Visconti

A Load Modeling Methodology for Steady State and Dynamic Simulations

Thesis presented to the Programa de Pós-graduação em Engenharia Elétrica of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Engenharia Elétrica. Approved by the undersigned Examination Committee.

Prof. Delberis Araujo Lima

Advisor

Departamento de Engenharia Elétrica – PUC-Rio

Prof. Antonio Padilha Feltrin

Universidade Estadual Paulista – UNESP

Prof. Djalma Mosqueira Falcão

Universidade Federal do Rio de Janeiro – UFRJ

Prof. José Luiz Rezende Pereira

Universidade Federal de Juiz de Fora – UFJF

Nelson Martins

Eletrobras Cepel

Rio de Janeiro, July 12th, 2019

All rights reserved.

Igor Ferreira Visconti

Graduated in Electrical Engineering at the Pontifical Catholic University (PUC-Rio) in 2001, and since 2002 works as a researcher in Electrical Energy Research Center (Eletrobras Cepel). Post-graduated (*latu-sensu*) in Analysis and Management of Software Projects at PUC-Rio in 2004, obtained his M. Sc. Degree also from PUC-Rio in 2010.

Bibliographic Data

Visconti, Igor Ferreira

A load modeling methodology for steady state and dynamic simulations / Igor Ferreira Visconti; advisor: Delberis Araujo Lima. – 2019.

132 f. : il. color. ; 30 cm

Tese (doutorado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica, 2019.

Inclui bibliografia

1. Engenharia Elétrica – Teses. 2. Modelagem de cargas. 3. Algoritmos genéticos. 4. Inferência estatística. 5. Conservação de energia através de redução de tensão. 6. Transitórios eletromecânicos. I. Lima, Delberis Araujo. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. III. Título.

CDD: 621.3

I dedicate this work to my daughter Iris, to my wife Roberta, and to my parents Angela e Jorge, without whom I would never have gotten so far.

Acknowledgement

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (Capes) – Finance Code 001.

I would like to thank all the people that helped me somehow in the years that I have studied at Pontifical Catholic University (PUC-Rio), including friends, classmates, teachers, employees, since I have spent half of my life in this institution, learning so many things that it would be impossible to describe them here, but many of them has changed my life forever and made me a more complete and prepared human being. In particular, I would like to thank my advisor Delberis Araujo Lima for the time he dedicated to this project.

I would like to thank my workmates at Cepel, for the contributions to clarify the questions that I could not find in the books. I would also like to thank them for the time that I was granted to study and develop better engineering skills to face the challenges that a research center presents all the time.

I would like to thank my family, for giving me the support, both emotional and financial, which gave me strength to move forward, even when I thought I could not do it. Jorge and Angela, my parents were so important and always present during my whole life, that I could not thank them enough, but at least I have tried in this few lines.

And, last but not least, I would like to thank my wife, my best friend, my partner in life, the mother of my daughter, Roberta.

Abstract

Visconti, Igor Ferreira; Lima, Delberis Araujo (advisor). **A Load Modeling Methodology for Steady State and Dynamic Simulations**. Rio de Janeiro, 2019. 132p. Tese de Doutorado - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

To simulate, predict and control Electric Power Systems (EPS), engineers need tools to model the components of this highly complex interconnected network. Many efforts over the past century were dedicated to develop mathematical models for generators, transmission lines, reactive power compensators, transformers and so on. The main components of the power systems are precisely represented by mathematical models, but the loads are still a source of uncertainty in the simulations, due to their random characteristics. It is well known that conservative load models super estimate power response to voltage deviations, and, on the other hand, over-optimistic load models may underestimate stability margins, leading a system to operate too close to its limit. It is necessary to establish load representations as close to reality as possible, in order to fully exploit grid resources. This work provides a methodology for load modeling, investigating and summarizing the steps of the process, whose can be implemented in a wide variety of ways. Data treatment, the choice of a load model representation and their parameters estimation are presented through real case studies, both for dynamic simulation and a steady state application. It is discussed how optimization and statistical inference concepts can be effective tools to reach better approximations on how load will respond to disturbances caused by voltage variations, whether these were spontaneous, due to control actions, or caused by short-circuits.

Keywords

Load Modeling; Genetic Algorithms; Statistical Inference; Conservation Voltage Reduction; Electromechanical Transients.

Resumo

Visconti, Igor Ferreira; Lima, Delberis Araujo. **Uma Metodologia de Modelagem de Cargas para Simulações em Regime Permanente e Dinâmicas**. Rio de Janeiro, 2019. 132p. Tese de Doutorado - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Para simular, prever e controlar os sistemas de energia elétrica, engenheiros precisam de ferramentas computacionais para modelar os componentes dessa rede interconectada altamente complexa. Muitos esforços ao longo do século passado foram dedicados a desenvolver modelos matemáticos para geradores, linhas de transmissão, compensadores de potência reativa, transformadores e assim por diante. Os principais componentes dos sistemas de potência são representados precisamente através de modelos matemáticos, mas as cargas ainda são uma fonte de incerteza nas simulações, devido à sua característica de aleatoriedade. Modelos de carga conservadores superestimam a resposta de potência a desvios de tensão, enquanto modelos de carga excessivamente otimistas podem subestimar as margens de estabilidade, deixando o sistema muito próximo do seu limite operacional. É preciso estabelecer representações de cargas tão próximas da realidade quanto possível, a fim de explorar os recursos de rede de modo mais eficiente. Este trabalho fornece uma metodologia para modelagem de carga, investigando e resumindo as etapas do processo, que podem ser implementadas de diversas maneiras. O tratamento de dados, a escolha de uma representação matemática do modelo de carga e sua estimação de parâmetros são apresentados através de estudos de caso reais, tanto para uma aplicação focada na dinâmica do sistema elétrica, quanto para uma aplicação em regime permanente. Discute-se como otimização e conceitos de inferência estatística podem ser ferramentas efetivas para alcançar melhores aproximações sobre como a carga responderá a perturbações causadas por variações de tensão, sejam estas variações espontâneas, ou devido a ações de controle, ou causadas por curtos-circuitos.

Palavras-chave

Modelagem de Cargas; Algoritmos Genéticos; Inferência Estatística; Conservação de Energia através de Redução de Tensão; Transitórios Eletromecânicos.

Table of Contents

1 Introduction	16
1.1. Motivation	18
1.2. Objectives and Contributions	19
1.3. Outline of the Thesis	20
1.4. Publications	21
2 Literature Review	23
2.1. Historical Perspective	24
2.2. Load Model Definition and Classification	25
2.2.1. Static Load Models	25
2.2.1.1. Constant Impedance Model (Z)	26
2.2.1.2. Constant Current Model (I)	27
2.2.1.3. Constant Power Model (P)	27
2.2.1.4. Polynomial Model (ZIP)	28
2.2.1.5. Exponential Model	29
2.2.2. Dynamic Load Models	30
2.2.2.1. Exponential Recovery Model	31
2.2.2.2. Induction Motor	34
2.2.3. Composite Load Models	36
2.2.4. Black-box Models	41
2.3. Load modeling approaches	42
2.3.1. Component-based load modeling	42
2.3.2. Measurement-based load modeling	43
2.4. Summary of the chapter	44
3 Measurement-based Load Modeling Methodology	46
3.1. System Identification Approach	46
3.1.1. System Identification concepts	47
3.2. Proposed Load Modeling Implementation	49
3.2.1. Data treatment	52

3.2.2. Model selection	52
3.2.3. Parameter Estimation	53
3.2.4. Model Validation	55
3.3. Summary of the chapter	55
 4 Load Modeling for Dynamic Simulation – A Case Study	 57
4.1. Case Study Description	58
4.2. Data treatment	59
4.3. Model Selection	63
4.3.1. Transfer Functions	65
4.4. Parameter Estimation	70
4.4.1. Genetic Algorithm	73
4.4.2. Higher order investigation and a comparison to induction motor different orders	75
4.4.3. Parameter estimation to describe a set of typical contingencies	78
4.4.4. Parameter estimation to reproduce one event of interest	83
4.5. Model Validation	85
4.6. Summary of the chapter	92
 5 Load Modeling for Conservation Voltage Reduction – A Case Study	 94
5.1. Case Study Description	96
5.2. Data treatment	104
5.2.1. Voltage Steps Identification Procedure	105
5.3. Model Selection	108
5.4. Parameter Estimation	110
5.4.1. Statistical Inference	112
5.4.1.1. Bootstrap method	116
5.5. Results	118
5.6. Summary of the Chapter	121
 6 Conclusions	 123
 7 References	 127

List of Figures

Fig. 1 – Summary of P deviations from the operating point P0, as a function of the different static load models described above	28
Fig. 2 – Power system stability definitions and classification [8]	31
Fig. 3 – Matlab Simulink implementation of the exponential recovery model (discrete-time formulation)	33
Fig. 4 – Demonstration of load response of the Exponential Recovery Load Model: after a voltage step (green line), at instant 30, that reduces voltage from 1.0 to 0.9 p.u., the load response (orange line) reduces from 1.0 p.u. to 0.8 p.u. temporally, and then reaches a steady state, around instant 40, remaining in 0.9 p.u. In this example, the load model's parameters set q is $T_p = 10$, $K_{pt}=2$ and $K_{pss}=1$.	33
Fig. 5 – Equivalent circuit of the induction motor load model	34
Fig. 6 – Schematic Diagram of the ZIP + Induction Motor	36
Fig. 7 – ADN equivalent diagram [15]	37
Fig. 8 – Back-to-back full converter model	38
Fig. 9 – Schematic diagrams of the system identification concepts: the diagram (a) shows the concept of voltage data as the input of models mp and mq, parameterized respectively by qp and qq, whose outputs are respectively the active power P(t) and the reactive power Q(t). The diagram (b) represents a models set M containing three different models (and their respective parameters sets) to represent the active power P(t) as a function of voltage input V(t).	49
Fig. 10 – Typical placement of disturbance recorders for distribution load model	49
Fig. 11 – Load modeling schematic procedure	51
Fig. 12 – A power electric grid diagram from a northeastern Brazilian utility	58
Fig. 13 – Case study measurement setup	59

Fig. 14 – Instantaneous voltages and currents recorded by the IED with a sampling rate of 1920 Hz	61
Fig. 15 – Positive sequence active (blue) and reactive (red) power components against positive sequence voltage (green) component.	61
Fig. 16 – Discrete Fourier results from snapshots of a pre-disturbance (green phasors) and “during the disturbance” (red phasors) voltages and currents shown in Fig. 15.	61
Fig. 17 – Processing data procedure	62
Fig. 18 – Parameters $w_{2p,q}$, $w_{2p,q}$ boundaries (f_1 and f_2 respectively) [26]	64
Fig. 19 – Simulink implementation of the discrete time transfer function of the proposed active power load model	66
Fig. 20 – Simulink implementation of the continuous time transfer function of the proposed active power load model	67
Fig. 21 – Analysis of homogeneous response from the system pole locations.	69
Fig. 22 – Estimating parameters of active power (first generation)	72
Fig. 23 – Estimating parameters of active power (fourth generation)	73
Fig. 24 – Estimating parameters of active power (after 200 generations)	73
Fig. 25 – Genetic Algorithm (GA) coding strategy for 2nd and 3rd order model.	74
Fig. 26 – 2nd order transfer function for reactive power load model	76
Fig. 27 – 3rd order transfer function for reactive power load model	76
Fig. 28 – 4th order transfer function for reactive power load model	77
Fig. 29 – 5th order transfer function for reactive power load model	77
Fig. 30 – Comparison between the measured P (lighter dashed curve), the P simulated by the estimated model (darker dotted curve) and the P described by a generic ZIP model	79

Fig. 31 – Validating the estimated active power model for the 2010 contingencies.	80
Fig. 32 – Comparison between the measured Q (lighter dashed curve), the Q simulated by the estimated model (darker dotted curve) and the Q described by a constant-impedance load model (solid curve).	81
Fig. 33 – Validating estimated reactive-power model for 2010 contingencies.	81
Fig. 34 – A single load model to represent reactive power for 24 different contingencies.	82
Fig. 35 – A single load model to represent active power for 24 different contingencies.	83
Fig. 36 – Reproducing a specific event: a distribution system containing DG and active power behavior.	84
Fig. 37 – Reproducing a specific event: a distribution system containing DG and reactive-power behavior.	85
Fig. 38 – Active power simulated at the 69 kV bus bar with DG by a constant-impedance load model (P_{zip}) and the last section's estimated model (P).	86
Fig. 39 – Reactive power simulated at the 69 kV bus bar with the constant-impedance (Q_{zip}) and estimated models (Q).	86
Fig. 40 – P_{med} is the active power measured, P_{calc} is the estimated load model, described by the equation at the top of the figure, and PZIP is 60% constant P and 40% constant Z (substation 1)	87
Fig. 41 – Comparison between PZIP and the estimated load model of Fig. 40 (substation 1) implemented in ANATEM.	87
Fig. 42 – Q_{med} is the reactive power measured, Q_{calc} is the estimated load model, described by the equation at the top of the figure, and QZIP is 100% constant Z (substation 1)	88
Fig. 43 – Comparison between QZIP and the estimated load model of Fig. 42 (substation 1) implemented in ANATEM.	88

Fig. 44 – Pmed is the active power measured, Pcalc is the estimated load model, described by the equation at the top of the figure, and PZIP is 60% constant P and 40% constant Z (substation 2).	89
Fig. 45 – Comparison between PZIP and the estimated load model of Fig. 44 (substation 2) implemented in ANATEM.	89
Fig. 46 – Qmed is the reactive power measured, Qcalc is the estimated load model, described by the equation at the top of the figure, and QZIP is 100% constant Z (substation 2).	90
Fig. 47 – Comparison between QZIP (constant Z) and the estimated load model of Fig. 46 (substation 2) implemented in ANATEM	90
Fig. 48 – Pmed is the active power measured, Pcalc is the estimated load model, described by the equation at the top of the figure, and PZIP is 60% constant P and 40% constant Z (substation 3)	91
Fig. 49 – Comparison between PZIP (60% constant P and 40% constant Z) and the estimated load model of Fig. 44 (substation 3) implemented in ANATEM.	91
Fig. 50 – Qmed is the reactive power measured, Qcalc is the estimated load model, described by the equation at the top of the figure, and QZIP is 100% constant Z (substation 3).	92
Fig. 51 – Comparison between QZIP (constant Z) and the estimated load model of Fig. 40 (substation 2) implemented in ANATEM	92
Fig. 52 – Schematic Diagram of primary substation monitoring	96
Fig. 53 – CLASS trials, extracted from [56]	97
Fig. 54 – Daily Profile of Avenham (Type 1), January/15 (Winter)	98
Fig. 55 – Daily Profile of Avenham (Type 1), July/14 (Summer)	99
Fig. 56 – Daily Profile of Fallowfield (Type 2), January/15 (Winter)	99
Fig. 57 – Daily Profile of Fallowfield (Type 2), May/15 (spring)	100
Fig. 58 – Daily Profile of Kitt Green (Type 3), January/15 (Winter)	100
Fig. 59 – Daily Profile of Kitt Green (Type 3), May/15 (Spring)	101
Fig. 60 – Daily Profiles registered in July/14 from one of the two parallel transformers in Kitt Green's substation: it is possible to	102

observe hours where the transformer is out of service (Monday, 21st), and also the hours where it was the only transformer operating the demand that was supposed to be fed by the two parallel transformers (Tuesday, 22th).

Fig. 61 – Diurnal patterns of load demand for a bell-shaped daily profile 103

Fig. 62 – Diurnal patterns of load demand for a residential daily profile 103

Fig. 63 – Slope calculation (green curves) to identify voltage steps 106

Fig. 64 – Example of the detection of a two-directional multi-step (OLTC successive tap changes reducing voltage a after a some time the OLTC sucessive tap changes boosting voltage) 107

Fig. 65 – The choice of the window size is not trivial 107

Fig. 66 – Voltage and active power data (upper plots) and their Savitzky-Golay smoothed counterparts below. 108

Fig. 67 – Step Response of the Exponential Recovery Load Model (5.4) 109

Fig. 68 – Active power load model parameters estimation for a summer event for the Golborne substation. 111

Fig. 69 – Reactive power load model parameters estimation for a summer event for the Golborne substation. 112

Fig. 70 – Histograms of 9 substations analyzed in the CLASS project. 116

Fig. 71 – Normal probability plots one the substations of Type 1, Type 2 and Type 3, presented at the first, second and third columns respectively. 117

Fig. 72 – An EDA approach to justify the bootstrap method to infer the expected value of CVR factor ass: the first column presents the normal probability plot of the ass distribution as it was estimated and the histogram of the same data distribution; the second column presents the same plot after the bootstrap resampling method. 118

Fig. 73 – CVR factors estimates' distributions 120

Fig. 74 – Scatter plot correlating CVR factor against loading conditions 121

List of tables

Table 1 – System Identification core concepts [22]	47
Table 2 – Second order transfer functions used for the proposed load modeling methodology	66
Table 3 – Summary of discrete and continuous time transfer functions to represent aggregated load	70
Table 4 – Group definitions for strata specification	115
Table 5 – CVR Factors expected values and their confidence intervals	119

1

Introduction

The electric power systems operation and planning always have been facing the challenge to balance the generation resources and load demand within the grid, in a reliable and cost-effective way. This work was motivated because load representations still need more accuracy to cover the most critical scenarios in power systems simulations, and they have to be updated due to a variety of new devices connected to the grid. In particular, distributed energy resources has been transforming the passive loads of the past into active distribution networks, which can even export their generation surplus to the transmission systems or become islanded from the major system during emergency events. These new distribution systems could be studied, and it should be investigated if it is possible to represent such systems with a single load model structure, with no physical meaning, using only data measurements taken from the frontier of the transmission system and these “active loads”. Time series analysis, big data and data mining are some knowledge areas that can help to improve load model estimation using field data. If in the past measurement-based load modeling was not possible to be implemented in a systematic way due to the lack of measurement devices installed in suitable points of the power grids, now it is possible to update continuously load model parameters, in order to evaluate dynamic equivalent of the system, using modern monitoring devices, fast communication links and high performance computing techniques.

Experts responsible for the operation of electric power systems rely their decisions on different static and dynamic simulations, aiming the optimization of their assets, the minimization of losses and the maximization of power transmission capacity. These goals should be achieved respecting the electric power system stability margins and investment considerations.

Electromechanical transient simulations analyze the behavior of systems after disturbances occurring in power grid. Many types of contingencies may cause sudden changes in the electrical network configuration, where the system leaves the

operating point pre-disturbance stable. Depending on the severity or place of origin of the occurrence, it is desirable to predict whether the system can reach another equilibrium point or will be unstable after the elimination of the disturbance by the protection system.

Load modeling has a great impact on system dynamics simulations, especially stability studies, where the dynamic reacceleration of industrial engines influences the restoring synchronism between the hydraulic generating units. Contingencies can provoke accelerations and decelerations in certain units or groups of generating units that are subjected to lose synchronization between them or in relation to the system.

Historically, mathematical models for generators, transmission lines, transformers and other elements that constitute and influence the behavior of electrical systems have been represented through sets of equations. However, load modeling for dynamic and static simulation remains a challenge, due to its stochastic behavior, and the complexity involved in lumping thousands of components that consume power from the grid in a single accurate representation. Besides the inherent randomness to predict and simulate load response to system perturbations, it is observed that the load demand varies with different time frames, which means there are expected cyclic variations during the 24 hours of a day, the days of the week, and also the seasons of the year.

This thesis provides the general guidelines to estimate a load representation using suitable data measurements from a point of common coupling between the system to be modeled, downstream the point of installation where the data measurement device is installed. It will present the steps involved in data selection, signal processing, system identification, parameter estimation and model validation. A flexible load model is proposed, providing an application case study. The same methodology is used to assess the potential to provide peak demand reduction and energy savings in another case study, that estimates load model representations that quantify how much energy is saved, and if it is possible to reduce demand during the peak hours.

1.1. Motivation

Optimistic load model can compromise system security, evaluating higher stability margins than the actual operating conditions impose; whereas the results of simulations that include pessimistic load models, reduces the power transfer capacity of transmission lines.

The response of the loads to voltage and frequency deviations of their nominal values is one of the factors that influence the loss or absence of system stability. Models that faithfully reproduce the loads behavior during and after disturbances increase the reliability of the results of simulations, enabling operators to anticipate potential emergency conditions or maximize power transmission of specific transmission lines.

In the last few years, distribution systems no longer can be defined as passive subsystems, supplied by the transmission system. With the increasing penetration of Distributed Generation (DG) connected to the grid through Medium or Low Voltage (MV or LV), the distribution systems with generating capacity have been denominated Active Distribution Networks (ADN), and this new concept has been demanded new mathematical representation and need more answers on how these subsystems will impact the overall system stability and system operation.

Also, with the advent of smart grids and due to environmental concerns, the network management has been increasingly demanding more efficiency, through the implementation of a modern infrastructure of monitoring, automation and bidirectional communications. One of the smart grid possible implementations is the old Conservation Voltage Reduction (CVR), which has become a cost effective solution to deferral network investments, due to the potential peak demand reductions, energy saving and other voltage control benefits, that depends on the different load characteristics of the subsystem to be modeled. Therefore, load modeling can quantitatively indicate how much energy can be saved by a voltage reduction, keeping it within its nominal acceptable range. It can indicate how much active power can be reduced during peak hours, by this scheduled and/or automatic voltage reduction.

In Brazil, the National System Operator (ONS) provides data from the integrated electric power system, comprising the transmission lines, generators and transformers, both for steady-state and stability software analysis. The whole Brazilian Integrated System is represented up to 69kV bars connected to all bus bars of this voltage level, static ZIP load models (section 2.2.1.4). The parameters for these load models are the same for many different loads, and they could be improved whenever there is suitable data measured in the point of common coupling between transmission and distribution systems, or the point between an upstream system and a downstream subsystem that it is desirable to be more accurately represented in power systems simulations. This thesis will discuss the way to do it.

1.2. Objectives and Contributions

This work intends to present the guidelines for developing load models to be used in electric power simulations, modeling the load response caused by voltage variations, using field-data.

This work will begin reviewing a collection of different mathematical representations of loads, reported in the scientific literature. Then, it will be detailed the guidelines to choose a suitable model structure and estimate the parameters set that best fits the available field data.

This comprehensive methodology consists in using the same optimization formulation (based on the minimization of the squared error between the measured and simulated outputs of the model), independent of the choice of the optimization technique or the mathematical structure, i.e., the load model. Different optimization algorithms were tested, but not extensively, only to show that the measurement-based load modeling guidelines are independent of the parameter estimation method, and also of the model choice. Nevertheless, the model specification can lead to different parameter set sizes, and therefore the optimization formulation may include different parameters constraints.

1.3. Outline of the Thesis

The thesis is organized as follows:

Chapter 2 will provide a literature review of the load definitions, the main mathematical structures to model the load behavior and two basic approaches regarding load modeling procedures: component-based load modeling and measurement-based load modeling.

Chapter 3 will provide the details of the methodology developed, inspired by the guidelines of System Identification aspects. It will be discussed a load model proposed by the author, that consists in representing an aggregated load connected to a bulk power system as a discrete time transfer function, and the parameters set constraints associated with this formulation, such as causal and stability issues of such a model characterization.

Chapter 4 will present the results from measurements obtained from a Brazilian transmission utility, and the complete process from selecting suitable data for load modeling, the signal data processing that filters and transforms the Multiple-Input Multiple-Output (MIMO) system into a Single-Input Single-Output (SISO) system; this is because stability system software usually is interested in positive sequence data. It will also be presented results from a simulation using a transient electromechanical software, using the complete Brazilian Interconnected System representation (up to 69kV) and some measurement-based dynamic load models, connected to 230/69kV, and its comparison to the adopted ZIP load models by ONS.

Chapter 5 will present another case study, using data measurements obtained in the distribution systems of Manchester-UK. These data were collected for a Conservation Voltage Reduction (CVR) investigation, and the measurement-based load modeling intends to quantify energy saving and peak demand reduction that could be achieved from different substations with different load profiles.

Chapter 6 will present a summary of the work, the conclusions and a summary of the guidelines to obtain more accurate representations of physical systems throughout mathematical representations that do not necessarily need to have physical meanings.

1.4. Publications

This research has produced several scientific articles that have been published in Brazilian and international symposiums:

1. Visconti, Igor F.; Lima, Delberis A.; Costa, Milanović, Jovica V., “Comprehensive analysis of Conservation Voltage Reduction: A real case study”, *2019 IEEE Milano PowerTech*, Italy.
2. Visconti, Igor F.; Lima, Delberis A.; Costa, J. M. C. de S. ; Sobrinho, N. R. de B. C., “Measurement-Based Load Modeling Using Transfer Functions for Dynamic Simulations”, *IEEE Transactions on Power Systems*, v. 29, p. 111-120, 2014.
3. Visconti, I. F.; Luiz Felipe Willcox ; Costa, J. M. S. C. ; Barros Sobrinho, N. R., “Metodologia para Estimaco de Modelos de Cargas a partir de Medies em Subestaes”, *XII Simpsio de Especialistas em Planejamento da Operao e Expanso Eltrica (SEPOPE)*, Rio de Janeiro, 2012.
4. Visconti, I. F.; Souza, L. F. W. ; Costa, J. M. S. C. ; Barros Sobrinho, N. R., “Measurement-based load modelling of systems with dispersed generation”, *Cigr Session 44*, 2012.
5. Rangel, R. D. ; Souza, L. F. W. ; Visconti, I. F. ; Macedo, N. J. P. ; Lima, M. C., “Validao de Modelos no ANATEM Atravs de Comparao com Sinais Reais de Medio”, *XXI SNPTEE - Seminrio Nacional de Produo e Transmisso de Energia Eltrica*, Florianpolis, 2011.
6. Costa, J. M. S. C. ; Barros Sobrinho, N. R. ; Souza, L. F. W. ; Visconti, I. F., “Estudos de Transitorios Eletromecnicos no Sistema de Transmisso da CHESF Utilizando Modelos de Carga Dinmica Baseados em Medies.” *XXI SNPTEE - Seminrio Nacional de Produo e Transmisso de Energia Eltrica*, Florianpolis 2011.
7. Costa, J. M. S. C. ; Barros Sobrinho, N. R. ; Souza, L. F. W. ; Visconti, I. F., “Representao De Consumidores Industriais Em Estudos De Transitorios Eletromecnicos.” *XIV ERIAC - XIV Encuentro Regional Iberoamericano del Cigr*, 2011, Ciudad del Este.

8. Visconti, I. F.; Souza, L. F. W. ; Costa, J. M. S. C. ; Barros Sobrinho, N. R., “Modelos De Carga Baseados Em Medições Para Simulações De Transitórios Eletromecânicos.” *XIV ERIAC - XIV Encuentro Regional Iberoamericano del Cigré, 2011, Ciudad del Este.*
9. Visconti, I. F.; Souza, L. F. W. ; Costa, J. M. S. C. ; Barros Sobrinho, N. R., “Estimação de Parâmetros de Modelos de Carga Dinâmicos usando Algoritmos Genéticos.” *III Simpósio Brasileiro de Sistemas Elétricos, Belém, 2010.*
10. Visconti, I. F.; De Souza, L. F. W. ; Costa, J. M. S. C. ; Sobrinho, N. R. B. C., “From power quality monitoring to transient stability analysis: Measurement-based load modeling for dynamic simulations”, *14th International Conference Harmonics Quality Power, Bergamo, Italy, Sep. 2010.*
11. Visconti, I. F.; De Souza, L. F. W., “Modelagem De Carga Baseada Em Medições Utilizando Algoritmos Genéticos” *XIII ERIAC - Décimo Terceiro Encontro Regional Iberoamericano do Cigré, Puerto Iguazu, Argentina, 2009.*

2 Literature Review

Due to the advent of new types of electronic loads, the increasing penetration of distributed generation and the modernization of wide area monitoring, the research on load modeling has received a renewed interest. The energy management systems will become high technology information centers, hopefully marching towards a more efficient asset utilization, assuming the mathematical model of the power system is accurate enough to predict its behavior and simulate scenarios evaluating accurate system response to disturbances and contingencies.

Although it is acknowledged the key aspect of representing loads accurately in dynamic power systems simulations, the most usual practice adopts static load models, and power system studies that lack of load composition may even ignore the influence of voltage variation in load response, assuming the classic constant power load model, as it is revealed in the survey reported in [2].

But it has been repeatedly reported that attempts to reproduce power systems large contingencies have failed, whenever it was used standard or very simplistic representations of subsystems defined as loads. One of the earliest reports of load modeling inadequacy dates from the Swedish blackout of 1983 since the simulated and recorded data did not fit properly [3].

This chapter presents in section 2.1 a brief historical chronology of some of the most remarkable publications in the load modeling research area. In section 2.2, it is presented a formal definition of the concepts and terminology used in this field, and an extensive list of mathematical representations of active and reactive power consumed by loads, as a function of voltage. The last section will present the two basic approaches for the load modeling development, which will be used whether in steady-state or in dynamic software simulations.

2.1. Historical Perspective

From the last 30 years, data recording and storage capabilities, as well as the number of devices deployed in power systems, have increased intensively. Load model development and the quality of model parameters have evolved, due to more suitable data available for the purpose of this task. In [2], the responses to the questionnaire of CIGRE's survey on load models have reported that "the power industry is taking full advantage of available monitoring systems for such purposes" (...) over 50% of cases load model parameters are being identified based on field measurements".

The first benchmark to review the load modeling is a paper published by the IEEE Task Force on Load Representation for Dynamic Performance [4] that has summarized the basic definitions on load model and provided guidelines to dynamic performance analysis in power system studies. A couple of years later, this Task Force published a bibliography containing several static and dynamic load models, altogether with their references [5]; a few months later, they published [6] "to promote better and advanced load modeling, and to facilitate data exchange among users of various production-grade simulation programs".

But apparently industry and system operators were not able to follow these guidelines for measurement-based load modeling, because data measurements devices were not often available and they were relatively more expensive. However, the modern power grids have increased their monitoring devices and automation infrastructure, towards smarter grids. Moreover, the increasing share of Distributed Generation (DG) connected to the grid through Medium or Low Voltage (MV or LV), transformed the distribution systems into Active Distribution Networks (ADN), and they ask for new mathematical representation to predict how these subsystems will impact the overall system stability and system operation.

2.2. Load Model Definition and Classification

This section aims to cover the most common used load models found in literature and in power systems simulation software. But first, it is important to formalize the concept of load from the viewpoint of power systems.

The term "load" takes on different meanings depending on the context in which it is used [4]. It could be:

- A device connected to the power system that consumes energy.
- The total energy consumed by all devices connected to the power system.
- A portion of the system which is not represented in detail, but it is treated as if it was a single power consumer element, connected to a bus.

In particular, the last of the above definitions establishes that, once chosen a load bus, all that is connected downstream of the bus is added in an equivalent load and this is measured in terms of power consumed by this downstream subsystem. The voltage measured at the load bus is defined as the disturbance variable or input variable of the model. There are load models that use the disturbance frequency and ambient temperature as disturbance variables [2].

Load models are traditionally classified into two main groups: *static load models* (section 2.2.1) and *dynamic load models* (section 2.2.2). There are other possibilities to cluster them into different groups, or even to combine them into *composite load models* (section 2.2.3). According to recommendations of the IEEE Task Force [4]-[6], the static load models would be used in steady-state simulations, while dynamic models would better represent load behavior in stability studies. It has been dedicated also a special section for the black box load models, whose do not take into account any physical meaning regarding its mathematical formulation; this approach was the choice for the case study of Chapter 4 of this thesis.

2.2.1. Static Load Models

If load response due to some voltage deviation reaches a steady state condition so fast that the measuring device is not able to detect the dynamics of the load change, it is considered that a static load model is sufficient to represent this phenomenon. Static load models express the active and reactive power at any

instant of time, depending on the magnitude of load bus voltage and the power system frequency at that moment [4], as it is formulated in (2.1).

$$\begin{aligned} P(t) &= f(V(t), freq(t)) \\ Q(t) &= g(V(t), freq(t)) \end{aligned} \quad (2.1)$$

When there are small or slow variations of voltage and/or frequency, the system returns to steady state quickly, and in such cases it is possible to model the static load models without loss of generality.

In the next subsections, the most common formulations of static models relating power and voltage variations are presented, using the variables above:

$P(t)$ is the active power, expressed in watts (W), as a function of the discrete time variable t .

$Q(t)$ is the reactive, expressed in “volt-amperes reactive” (var), as a function of the discrete time variable t .

$V(t)$ is the voltage expressed in volts (V), as a function of the discrete time variable t .

V_0, P_0 and Q_0 are respectively the values of the voltage active and reactive powers in steady state before the disturbance or, mathematically speaking, the system equilibrium point, *i.e.* $P_0 = P(t=0)$, $Q_0 = Q(t=0)$ and $V_0 = V(t=0)$.

2.2.1.1. Constant Impedance Model (Z)

The constant impedance model expresses variations in power proportionally to the square of the voltage deviations, as it is shown in (2.2):

$$\begin{aligned} P(t) &= P_0 \left(\frac{V(t)}{V_0} \right)^2 \\ Q(t) &= Q_0 \left(\frac{V(t)}{V_0} \right)^2 \end{aligned} \quad (2.2)$$

Whenever short-circuits occur upstream the load bus, the latter will measure a voltage sag, and if the characteristics of the downstream subsystem behaves like a constant impedance load type, there will be a current reduction in the transmission and distribution lines.

The Brazilian national system operator adopts this load model for static simulations (power flow) and dynamic (transient electromechanical) for the reactive power, and also to model the active power when the ratio $V(t)/V_0 < 0.7$ occurs during one disturbance simulation in the system. In several articles found in the literature, including [6], this is considered a too simplistic practice.

2.2.1.2. Constant Current Model (I)

The constant current model expresses variations in power proportionally to voltage deviations, as it is shown in (2.3):

$$\begin{aligned} P(t) &= P_0 \left(\frac{V(t)}{V_0} \right) \\ Q(t) &= Q_0 \left(\frac{V(t)}{V_0} \right) \end{aligned} \tag{2.3}$$

Whenever short-circuits occur upstream the load bus, the latter will measure a voltage sag, and if the characteristics of the downstream subsystem behaves like a constant current load type, there will be no significant current variation in the transmission and distribution lines.

2.2.1.3. Constant Power Model (P)

This model describes an idealized indifferent relationship between voltage variations and power variations measured from a load bus:

$$\begin{aligned} P(t) &= P_0 \\ Q(t) &= Q_0 \end{aligned} \tag{2.4}$$

If a short-circuit upstream the load bus occurs, the latter will measure a voltage sag, and if the characteristics of the downstream subsystem behaves like a constant power load type, there will be a current increase in the transmission and distribution lines, since the power remains close to the pre-disturbance level.

This model is considered the most conservative representation of a load system in power system simulations, and therefore it is used in many studies whose results underestimate system capacity.

2.2.1.4. Polynomial Model (ZIP)

Probably the most popular load model used in steady state and dynamic simulations, it consists in a second order polynomial relationship between power and voltage deviations.

$$\begin{aligned} P(t) &= P_0 \left[p_z \left(\frac{V(t)}{V_0} \right)^2 + p_l \left(\frac{V(t)}{V_0} \right) + p_p \right] \\ Q(t) &= Q_0 \left[q_z \left(\frac{V(t)}{V_0} \right)^2 + q_l \left(\frac{V(t)}{V_0} \right) + q_p \right] \end{aligned} \quad (2.5)$$

This model is a linear combination of the constant impedance, the constant current and the constant power models, and its parameters set is subject to the linear constraint in 2.6:

$$\begin{aligned} p_z + p_l + p_p &= 1 \\ q_z + q_l + q_p &= 1 \end{aligned} \quad (2.6)$$

The parameters sets to be estimated are $\theta_p = [p_z, p_l]$ and $\theta_q = [q_z, q_l]$, because the parameters p_p and q_p are obtained due to the constraint (2.6). If it is desired to keep the physical meaning of the load model, the parameters domain should be range between 0 and 1, and most simulation software only allows these limits, but theoretically any real numbers that satisfy (2.6) could be accepted.

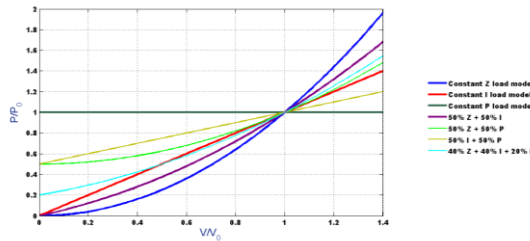


Fig. 1 - Summary of P deviations from the operating point P_0 , as a function of the different static load models described above

It is also known in the literature, but rarely used in practice, the frequency deviations included in the ZIP representation, as it is formulated in (2.7):

$$\begin{aligned} P(t) &= P_0 \left[p_z \left(\frac{V(t)}{V_0} \right)^2 + p_l \left(\frac{V(t)}{V_0} \right) + p_p \right] [1 + K_{pf} \Delta_{freq}] \\ Q(t) &= Q_0 \left[q_z \left(\frac{V(t)}{V_0} \right)^2 + q_l \left(\frac{V(t)}{V_0} \right) + q_p \right] [1 + K_{qf} \Delta_{freq}] \end{aligned} \quad (2.7)$$

The parameters sets to be estimated are $\theta_p = [p_z, p_l, K_{pf}]$ and $\theta_q = [q_z, q_l, K_{qf}]$. K_{pf} and K_{qf} are additional parameters to the ZIP parameters set, that quantify how frequency deviations from the rated value do affect the load, assuming a linear relationship. In most cases, the frequency-dependent model is ignored, but it may be required for describing the effects on damping of oscillations.

2.2.1.5. Exponential Model

Exponential load model is a static load model that describes a nonlinear relationship between power and voltage. In some cases, the dependency between reactive power and voltage is more than quadratic, when the ZIP model “loses physical meaning”.

$$\begin{aligned} P(t) &= P_0 \left(\frac{V(t)}{V_0} \right)^{\alpha_p} \\ Q(t) &= Q_0 \left(\frac{V(t)}{V_0} \right)^{\alpha_q} \end{aligned} \quad (2.8)$$

The parameters sets to be estimated are $\theta_p = [\alpha_p]$ and $\theta_q = [\alpha_q]$. In [7], it is described the equivalence between Exponential and ZIP load models' parameters set, based on Taylor series, restricted to small voltage changes. The equations (2.9), uses the parameters defined in (2.6) and (2.8):

$$\begin{aligned} \alpha_p &\approx \frac{2 * p_z + 1 * p_l + 0 * p_p}{p_z + p_l + p_p} \\ \alpha_q &\approx \frac{2 * q_z + 1 * q_l + 0 * q_p}{q_z + q_l + q_p} \end{aligned} \quad (2.9)$$

To assess the exponential parameter through active power and voltage data, it is straightforward:

$$P(t) = P_0 \left(\frac{V(t)}{V_0} \right)^{\alpha_p} \Rightarrow \ln \left(\frac{P(t)}{P_0} \right) + \alpha_p \ln \left(\frac{V(t)}{V_0} \right) \Rightarrow \alpha_p \approx \frac{\frac{P - P_0}{P_0}}{\frac{V - V_0}{V_0}} \quad (2.10)$$

or

$$P(t) = P_0 \left(\frac{V(t)}{V_0} \right)^{\alpha_p} \Rightarrow \log \left(\frac{P(t)}{P_0} \right) + \alpha_p \log \left(\frac{V(t)}{V_0} \right) \Rightarrow \alpha_p = \frac{\log \frac{P}{P_0}}{\log \frac{V}{V_0}}$$

It is also a suitable estimator of the CVR factor that quantifies peak demand reduction, as it will be discussed in Chapter 5.

2.2.2. Dynamic Load Models

Dynamic response of loads is a key aspect to stability analysis, particularly to represent slow voltage recovery after large system disturbances. A dynamic load model aims to describe the load response over time, and relates its output at a given instant as a function of the input at the same instant and also the output and input from preceding instants.

$$\begin{aligned} P(t) &= f(V(t), freq(t), t) \\ Q(t) &= g(V(t), freq(t), t) \end{aligned} \quad (2.11)$$

Dynamic load models may be formulated as a set of differential equations in time-continuous domain or a set of differences equations in discrete-time domain. Dynamic load models, or dynamic equivalents of grid subsystems, are commonly used in power system stability simulations, but this concept is rather wide and therefore in [8], definitions and classification were established and Fig. 2 presents the classification of the subject, based on:

- the physical nature of the mode of instability;
- the intensity of disturbance;
- the time span necessary to perform the analysis of the results, whether from simulation or field measurements.

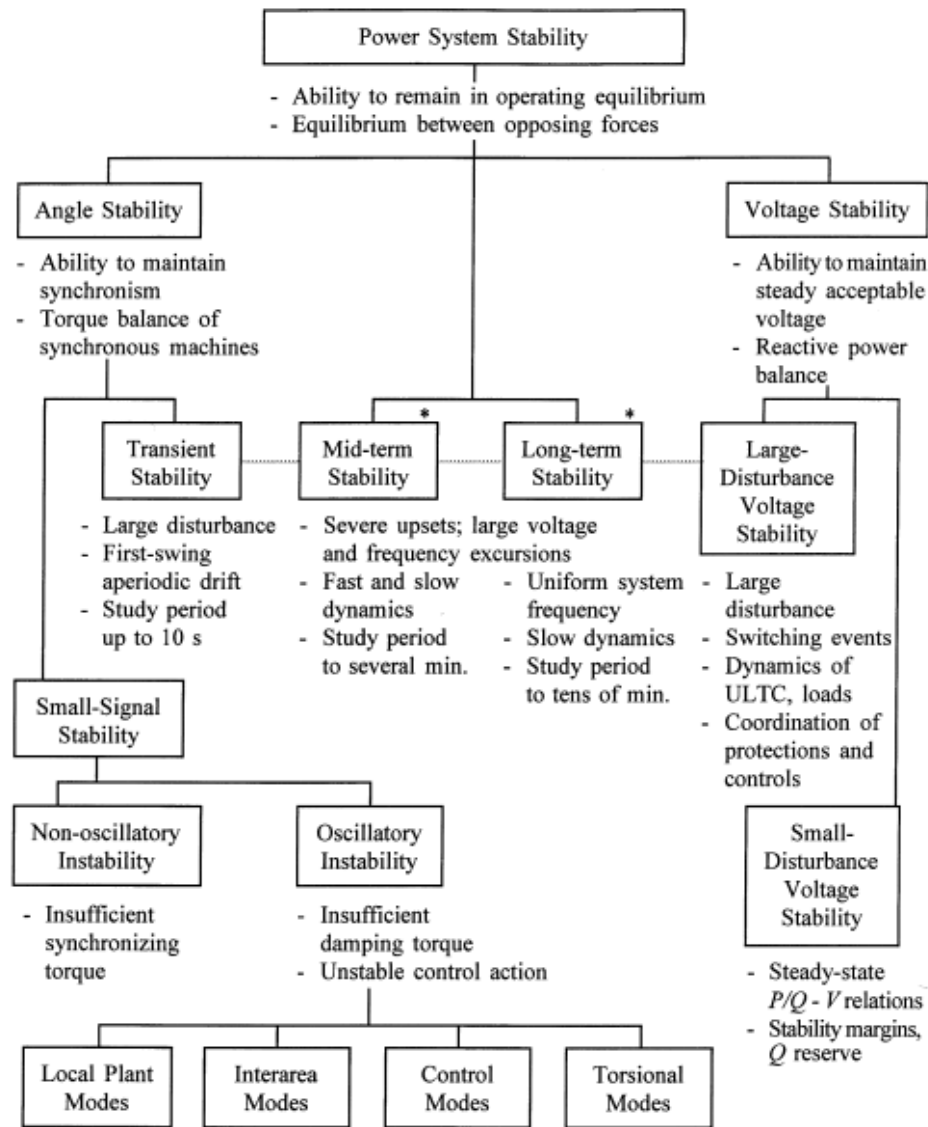


Fig. 2 - Power system stability definitions and classification [8]

Next, it will be presented a list of models that takes into account the past samples of both voltage and power variations, to describe load behavior during and after voltage disturbances.

2.2.2.1. Exponential Recovery Model

This load model was proposed by David J. Hill in [9] and it is basically a combination of two exponential models (2.8), while one of the exponential load models is associated with the transitory response, the other one is associated with

the steady state response, including a time constant that models the recovery of the power to the steady state.

According to the David Hill's conclusions in the seminal paper, this load model was motivated by the observed non-linearity from the measurements of power supplied to induction motors, heating loads and tap changers after voltage steps. Furthermore, this paper has been cited by more than a hundred papers related to load modeling, and it consists of a simple choice to include dynamics within a mathematical representation, and the forced response to a non-linear model. The formulation of this model is shown in:

$$T_p \frac{dP_r(t)}{dt} + P_r(t) = P_0 \left(\frac{V(t)}{V_0} \right)^{K_{ps}} - P_0 \left(\frac{V(t)}{V_0} \right)^{K_{pt}} \quad (2.12)$$

$$P(t) = P_r(t) + P_0 \left(\frac{V(t)}{V_0} \right)^{K_{pt}}$$

Where:

- $P_r(t)$: Real power recovery.
- P_0, V_0 : Pre-disturbance values of real power and voltage (before the voltage change), respectively.
- T_p : Recovery time constant.
- K_{ps} : Steady-state real power voltage exponent.
- K_{pt} : Transient real power voltage exponent.
- $P(t)$: Real power consumption of the load.

Fig. 3 presents an implementation of (2.12) in the discrete-time domain, within the Matlab/Simulink development environment. The choice to implement in the Simulink the model's first order differential equation (2.12) as a first order difference equation is because, in practice, data measurements are always recorded as discrete samples of voltage and current signals, thus the formulation must be adapted to treat the data measurements input and evaluate the simulated output. This load model will be used in the parameter estimation investigation of the case study described in Chapter 5; the data available from the substations were recorded with a sampling rate of 1 Hz, so the data set consists in discrete output/input samples of a continuous time physical system.

The two exponential load models of (2.12) were implemented as encapsulated subsystems in the schematics of Fig. 3. The differential operand of (2.12) was substituted in the Simulink environment by the Discrete Time Integrator. Simulink Design Optimization package provides many parameters estimation tools and it was a helpful tool to perform a load modeling investigation following the guidelines of the system identification approach. The parameters set to be estimated is $\theta_p = [T_p, K_{pss}, K_{pt}]$.

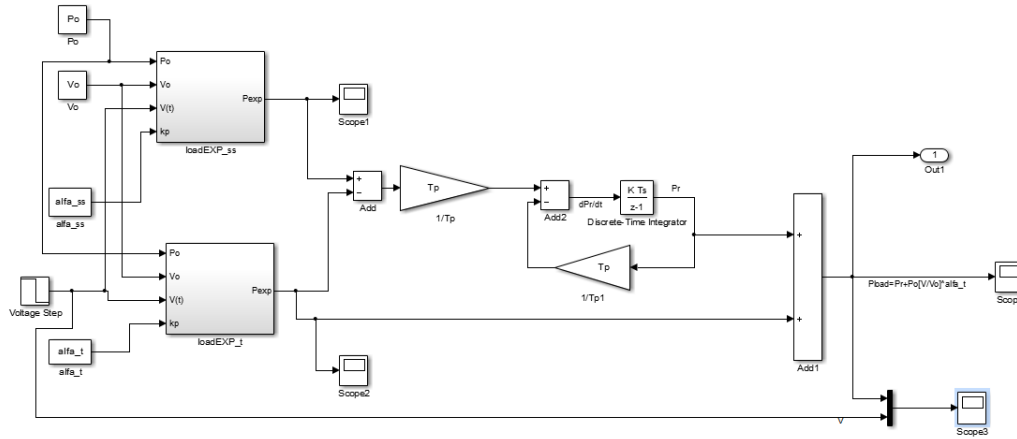


Fig. 3 - Matlab Simulink implementation of the exponential recovery model (discrete-time formulation)

Fig. 4 presents the input and output signals of the Simulink implementation, simulating a voltage step (green line) as the input of the exponential recovery model, and the corresponding active power (orange line) resulted from the voltage step and the parameters set configured in this simulation. It illustrates the transient response (more severe) and the steady state load response.

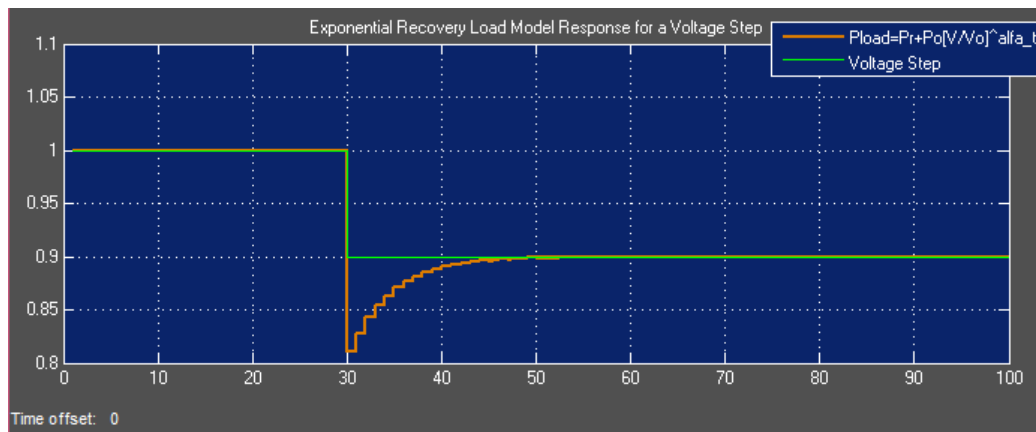


Fig. 4 – Demonstration of load response of the Exponential Recovery Load Model: after a voltage step (green line), at instant 30, that reduces voltage from 1.0 to 0.9 p.u., the load response (orange line) reduces from 1.0 p.u. to 0.8 p.u. temporally, and then reaches a steady state, around instant 40, remaining in 0.9 p.u. In this example, the load model's parameters set θ is $T_p = 10$, $K_{pt}=2$ and $K_{pss}=1$.

2.2.2.2. Induction Motor

It is the most common representation for modeling dynamic loads, assuming that induction motors are the load devices that predominate in electric power systems, with estimated percentages up to 70% of the total energy consumed by the load buses.

Most power electric dynamic simulation software contain induction motor models for representing dynamic loads, for the user to set tunable parameters, such as the stator impedance parameters (R_s , and X_s), the rotor (R_r and X_r) and the magnetizing reactance (X_m) of an equivalent induction motor (Fig. 5).

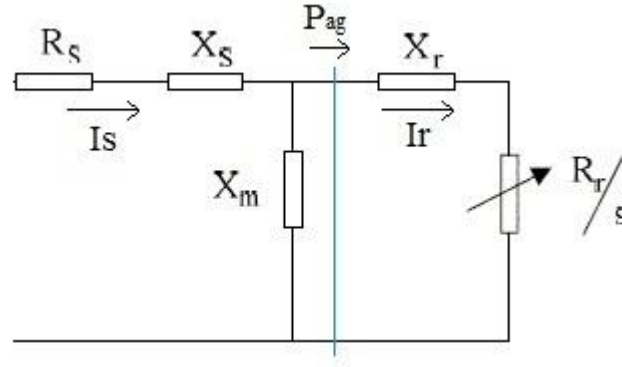


Fig. 5 – Equivalent circuit of the induction motor load model

The following differential equations describe the third order induction motor load model:

$$\frac{\partial^2 \theta}{\partial t^2} = \frac{\partial \omega}{\partial t} = -\frac{1}{2H} [T_m - T_e] = \frac{\omega_s - \omega_r}{\omega_s} \quad (2.13)$$

$$\frac{\partial \omega}{\partial t} = -\frac{1}{2H} [(A\omega^2 + B\omega + C)T_0 - (E'_d I_d + E'_q I_q)]$$

$$\frac{\partial E'_q}{\partial t} = -\frac{1}{T'} [E'_q + (X - X')I_d] + (\omega - 1)E'_d \quad (2.14)$$

$$\frac{\partial E'_d}{\partial t} = -\frac{1}{T'} [E'_d + (X - X')I_q] - (\omega - 1)E'_q$$

$$I_d = \frac{1}{(R^2 + X^2)} [(U_d - E'_d)R_s + X'(U_q - E'_q)] \quad (2.15)$$

$$I_q = \frac{1}{(R^2 + X^2)} [(U_q - E'_q)R_s + X'(U_d - E'_d)]$$

$$T' = \frac{X_r + X_m}{R_r} \quad (2.16)$$

$$X = X_s + X_m \quad (2.17)$$

$$X' = X_s + \frac{X_m X_r}{X_m + X_r} \quad (2.18)$$

$$A + B + C = 1 \quad (2.19)$$

$$P_{ag} = \frac{R_r}{s} I_r^2 = P_{sh} + P_{loss} \Rightarrow P_{sh} = \frac{R_r}{s} I_r^2 - R_r I_r^2 \quad (2.20)$$

$$P_{mot} = U_d I_d + U_q I_q$$

$$Q_{mot} = U_d I_d - U_q I_q \quad (2.21)$$

Where:

s = "slip", ratio between rotor speed and synchronous speed ω_r / ω_s (p.u.).

θ_r —angular position of the rotor regarding the synchronous reference (rad).

ω_s – The synchronous angular speed (rad/s).

ω_r – rotor angular speed (rad/s).

R_s e X_s – Stator winding resistance and reactance (p.u.).

R_r e X_r – Rotor winding resistance and reactance (p.u.).

X_m – Motor magnetizing reactance (p.u.).

H – Motor rotor inertia constant.

A – Coefficient of torque proportional to the square of ω .

B – Coefficient of torque directly proportional to ω .

C – Coefficient of constant torque.

T_0 – Torque in steady state.

T' – Transient open circuit time constant (rad).

I_d e I_q – Direct and quadrature axis stator currents.

U_d e U_q – Direct and quadrature axis stator voltages.

E'_d e E'_q – Transient direct-axis and quadrature-axis electromotive force.

X' – Transient reactance of the induction motor.

P_{ag} – Active power transferred by the gap to the rotor.

P_{sh} – Mechanical power transferred to the shaft.

The equations (2.13) are identical, describing the acceleration ratio of the rotor due to the imbalance between the electromagnetic and a mechanical torques of the motor. The equations (2.14) and (2.15) represent the dynamics of the rotor circuit, which express the decay of the magnetic flux of the rotor. The equation in (2.16) characterize the decay of the transients of the rotor when the stator is at a vacuum. The equation (2.20) represents the active power transferred by the rotor gap and its relation to the mechanical power transferred to the axis and the loss in rotor resistance, while (2.21) presents the active and reactive power models P_{mot} and Q_{mot} as a function of the direct-axis and quadrature.

It is necessary to use numerical integration methods to solve these ordinary nonlinear differential equations, such as Euler or Runge-Kutta. The parameters set to be estimated is $\theta_p = [R_s, X_s, X_m, R_r, X_r, H, A, B]$.

2.2.3. Composite Load Models

Composite load models combine both static and dynamic load components into a single aggregated load model. In [10], the authors present results for parameter estimation of the following combinations:

- ZIP + Induction Motor
- ZIP + difference equations
- Exponential + Induction Motor
- Exponential + difference equations

Undoubtedly, the most common is the ZIP + IM, depicted in Fig. 6, used in [12]-[14].

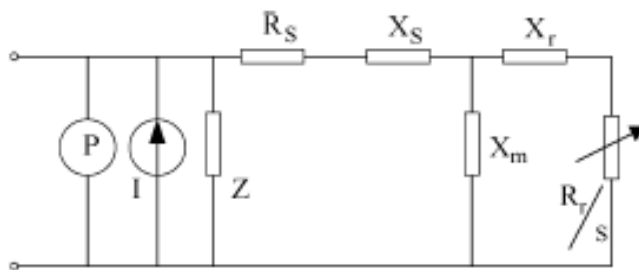


Fig. 6 - Schematic Diagram of the ZIP + Induction Motor

This combination of the equations from sections 2.2.1.4 and 2.2.2.2, introduces an extra parameter, that express the percentage K_{pm} of the contribution of the induction motor component to the load model, and naturally the contribution of the static part would be $1 - K_{pm}$.

In [15], the authors proposed to represent an Active Distribution Networks (ADN), based on a combination of different load model components, formulating it as a seventh-order nonlinear state space format. Fig. 7 illustrates this *grey-box* approach, that combines both a “physical meaning” modeling and a purely mathematical modeling.

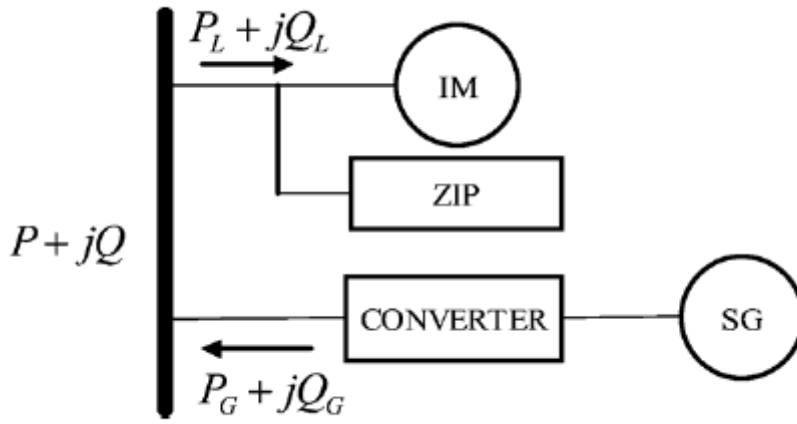


Fig. 7 – ADN equivalent diagram [15]

The load is represented by the $P_L + jQ_L$ flow, where its corresponding composite load model ZIP+IM diagram is already depicted in Fig. 6, using the article's notation:

$$\frac{\partial E'_m}{\partial t} = \frac{1}{T'_{dm}} \left[-\frac{X_m}{X'_m} E'_m + \left(\frac{X_m - X'_m}{X'_m} \right) V \cos \delta_m \right] \quad (2.22)$$

$$\frac{\partial \delta_m}{\partial t^2} = \omega_m - \omega_s - \left(\frac{X_m - X'_m}{X'_m} \right) \frac{V}{T'_{dm} E'_m} \sin \delta_m \quad (2.23)$$

$$\frac{\partial \omega_m}{\partial t} = -\frac{1}{H_m} \left(\frac{E'_m V}{X'_m} \sin \delta_m + T_m \right) \quad (2.24)$$

$$P_L = P_{ZIP_0} \left[p_Z \left(\frac{V}{V_0} \right)^2 + p_I \left(\frac{V}{V_0} \right) + p_P \right] - \left(\frac{V}{X'_m} E'_m \sin \delta_m \right) \quad (2.25)$$

$$Q_L = Q_{ZIP_0} \left[q_Z \left(\frac{V}{V_0} \right)^2 + q_I \left(\frac{V}{V_0} \right) + q_P \right] - \left(\frac{V^2}{X'_m} - \frac{V}{X'_m} E'_m \cos \delta_m \right) \quad (2.26)$$

The converter-connected generator is composed of a third order synchronous generator model that interfaces with the grid via a back-to-back full converter model [16], [17]. The real power flow through the converter is balanced via the DC-link (the capacitor linking between inverter and rectifier), as it can be seen in the diagram of Fig. 8.

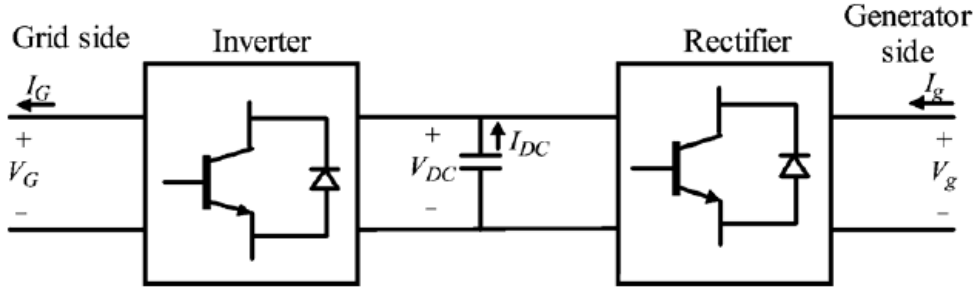


Fig. 8 - Back-to-back full converter model

The dynamic parts of the converter-connected generator can be described by:

$$\begin{aligned} \frac{\partial E'_g}{\partial t} &= -\frac{1}{T'_{dg}} [E_{FD} - E'_g - (X_g - X'_g)I_d] \\ &= -\frac{1}{T'_{dg}} \left[E_{FD} - E'_g \frac{X_g}{X'_g} + \left(\frac{X_g - X'_g}{X'_g} \right) V \cos \delta_g \right] \end{aligned} \quad (2.27)$$

$$\frac{\partial \omega_g}{\partial t} = \frac{1}{H_g} (T_m - T_e - D\omega_g) \quad (2.28)$$

$$\frac{\partial \delta_g}{\partial t} = \omega_g \quad (2.29)$$

$$\frac{\partial V_{DC}}{\partial t} = \frac{1}{CV_{DC}} (V_{dg}I_{dg} + V_{qg}I_{qg} - V_{DG}I_{DG} - V_{QG}I_{QG}) \quad (2.30)$$

$$P_G = \left(\frac{V}{X'_g} E'_g \sin \delta_g \right) + V_{DC}I_{DC} \quad (2.31)$$

$$Q_G = \left(\frac{V}{X'_g} E'_g \cos \delta_g - \frac{V^2}{X'_g} \right) + K_q V_{DC}I_{DC} \quad (2.32)$$

The generic nonlinear state space model of the system is given by

$$\begin{aligned}\frac{dx}{dt} &= Ax + Bu + f(x) \\ y &= Cx + Du = g(x)\end{aligned}\tag{2.33}$$

Then, the input, output and state vector are defined as:

$$\begin{aligned}x &= [E'_m, \delta_m, \omega_m, E'_g, \omega_g, \delta_g, V_{DC}]^T \\ u &= [V, \omega_s] \\ y &= [P_G - P_L, Q_G - Q_L]\end{aligned}\tag{2.34}$$

And the state space matrixes are:

$$A = \begin{bmatrix} \frac{-X_m}{T'_{dm}X'_m} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \frac{-V}{H_mX'_m} \sin \delta_m & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{-X_g}{T'_{dg}X'_g} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{-V}{H_gX'_g} \sin \delta_g & \frac{-D}{H_g} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{K_1}{CV_{DC}} \end{bmatrix}\tag{2.35}$$

$$B = \begin{bmatrix} \frac{X'_m - X_m}{T'_{dm} X'_m} \cos \delta_m & 0 \\ \frac{X'_m - X_m}{T'_m E'_m X'_m} \sin \delta_m & -1 \\ 0 & 0 \\ \frac{X_g - X'_g}{T'_{dg} X'_g} \cos \delta_m & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad (2.36)$$

$$C = \begin{bmatrix} \frac{V}{X'_m} \sin \delta_m & 0 & 0 & \frac{V}{X'_g} \sin \delta_g & 0 & 0 & I_{DC} \\ \frac{V}{X'_m} \cos \delta_m & 0 & 0 & \frac{V}{X'_g} \cos \delta_g & 0 & 0 & K_q I_{DC} \end{bmatrix} \quad (2.37)$$

$$D = \begin{bmatrix} -P_{zIP_0} \left(p_z \frac{V}{V_0^2} + p_l \frac{1}{V_0} \right) & 0 \\ -Q_{zIP_0} \left(q_z \frac{V}{V_0^2} + q_l \frac{1}{V_0} \right) - \frac{V}{X'_g} - \frac{V}{X'_m} & 0 \end{bmatrix} \quad (2.38)$$

$$f(x) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{E_{FD}}{T'_{dg}} \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (2.39)$$

$$g(x) = \begin{bmatrix} -P_{zIP_0}(p_P) \\ -Q_{zIP_0}(q_P) \end{bmatrix} \quad (2.40)$$

2.2.4. Black-box Models

Any load model that is not constructed according to physical laws can be classified as black-box models. State space representation of dynamic systems, differential and difference equations also can be included in this definition.

In the bibliography of load models published in [5], there are a few black-box formulations. From this reference, it was extracted some examples of dynamic load models without physical meaning, which is another way to say that they are purely mathematical expressions. In (2.22), using the same notation presented in the original reference [18], this load model was called by “Computer-based On Line Model”

$$P(t) = \sum_{i=1}^{np} a_i P(t-i) + \sum_{i=0}^{nv} b_i V(t-i) + \sum_{i=0}^{nf} c_i freq(t-i) + e(t) \quad (2.41)$$

Where:

- t is the discrete time index.
- $P(t-i)$ are observable load response outputs, expressed in p.u.
- $V(t-i)$ are observable voltage inputs, expressed in p.u.
- $freq(t-i)$ are observable frequency inputs, expressed in p.u.
- $e(t)$ are unobservable output disturbances, a zero mean sequence of Gaussian variables.
- a_i, b_i, c_i are regression coefficients of P, V and $freq$
- np, nv and nf are the regression orders of P, V and $freq$.

According to the system identification core definitions, that will be presented later in section 3.1 (Table 1), a model set $M = \{m_k\}$ could be composed of different models characterized by different regression coefficients and orders, defining a search space of parameterized model structures. These elements could also be parameters in a system identification procedure [21], [36].

$$m_k = [np_k, nv_k, nf_k] \quad (2.42)$$

Where k is an index to identify an item among the model set. The total numbers of parameters for each m_k is $np_k + nv_k + nf_k$, and the parameter set to be estimated is:

$$\theta_p(m_k) = [a_1, a_2, \dots, a_{np_k}, b_1, b_2, \dots, b_{nv_k}, c_1, c_2, \dots, c_{nf_k}] \quad (2.43)$$

This the most general formulation of discrete time domain dynamic systems (with two inputs). Using the Z transform [19], the equation (2.41) can be express as discrete time transfer functions, and manipulating it using Tustin, Forward or Backward transformation, it can be represented as a continuous time transfer function. In section 3.2, this equivalence will be developed.

2.3. Load modeling approaches

Since the cost of installing measuring devices have shrank in the last decades, and the data storage capabilities have increased greatly, it became more feasible to estimate more accurate load representations, improving the quality and upgrading load behavior for the different power systems simulations.

2.3.1. Component-based load modeling

The *component* or *knowledge-based approach* builds load model upon different load classes (such as residential, industrial or commercial classes), assuming that each class has similar load compositions and characteristics. It is also needed to assess the percentages of each class that comprises the aggregated load model, which is the whole demand supplied to a load bus bar. And last, but not least, the typical characteristic of each load class, or component, determined through a set of parameters, relating P and Q with respect to voltage magnitude and frequency. Typical components that may be assumed to compose residential load classes are air-conditioning, electronics, lamps and electric showers, whereas in industrial load classes the small and large induction motors are the most representative components. Each of these components may have different model structures, or can share the same structure with different parameters.

This approach is not able to take into account seasonality, or random weather variations uncorrelated with season, unless several scenarios are studied to estimate different load component characteristics, or variation in the percentages of components comprised in a load class. It has become more obvious that load

composition varies from one substation to another, and the simplification in assuming that all load classes can be represented in the same way, with the same components and the same percentage is likely to fail to represent load behavior with accuracy.

Once this component-based approach is built upon survey data from end use consumers, or previous assumptions based on other information, such as social/economical characteristics of the area, it is pretty obvious that it is necessary to update the percentages of components within the load class, and the parameters from the components characteristics that will vary over the hours, days and months.

2.3.2. Measurement-based load modeling

Obtaining an accurate load model, whose mathematical structure is based on physical laws may be important to gain knowledge and insight into the behavior of the load response, associating parameters to specific parts of the real systems, but the complexity of the aggregation of an enormous quantity of electric devices consuming energy varying with time seems to be not feasible to identify such a physical model (white box [21]). System identification methods are the best choice for estimation of unknown parameters from more flexible mathematical structures that does not need necessarily physical interpretation.

The main idea of measurement-based load modeling is to select typical input-output system data, which means select active and reactive power data (outputs), and voltage data (input), disturbed from steady state by sags caused by short-circuits originated outside the distribution system. The next step is to estimate a set of parameters for a transfer function that better fits these input-output data. It is desired a structure as simple as possible, with a minimum number of parameters, that best fits measured data that expresses real load behavior.

If load modeling main goal is to develop a generic load model, aiming to represent most typical disturbances used in stability simulation studies, load model parameter estimation needs a representative set of voltage data, and their associated active and reactive power data. Selecting this data properly is a key factor of this task. It is not feasible to achieve a model representation that simulates any

occurrence, at any operational state, but it is possible to cluster different operational conditions to estimate separately different set of parameters for each cluster of data.

A different approach is when load modeling main goal aims to reproduce, for example, a particular occurrence that has impacted the system within electric power simulation software. In this case, the best register for parameter estimation is the voltage sag that caused active and reactive power deviations from its initial operational point. All data set shall be recorded for a time long enough to cover from system pre-disturbance state until post-fault steady state has been reached. Fig. 2 specifies time frames for different purposes, consequently indicates the amount of post fault data to be used for each of the particular phenomena.

2.4.

Summary of the chapter

In this chapter it was presented a literature review regarding load modeling concepts and definitions. It was also presented the most used load models classified by their static, dynamic or a combination of both. In [2], the international survey found out that the most used load models are the ZIP, presented in 2.2.1.4, and the exponential, presented in 2.2.1.5; according to the same referenced survey, for steady state analysis, the constant power, presented in 2.2.1.3, is the most widely used; for dynamic studies, the exponential was reported to be the most used, but the induction motor load model representation, presented in 2.2.2.2, is also commonly used, with or without a static ZIP in parallel, characterizing a composite load model representation, as described in section 2.2.3.

The choice of the load model structure could be made *a priori*, or *a posteriori* [10], [21]. The industry acceptance or the availability in established simulation software may also guide the *a priori* choice of the load model structure.

For measurement-based load modeling, some assumptions can be taken to achieve a representation good enough for the purpose of your study. A good quality model must be able to fit data measurements accurately, but it also should not be too complex, for example with a number of parameters too big. This is the reason why the induction load motor is rarely used above its third order representation, thus neglecting some dynamics that will not influence much the equivalent representation. It is also acceptable to use linearized models that can facilitate

parameter estimation, and model implementation. In the next chapter, the measurement-based approach will be described in more details, and how the proposed methodology was built upon the system identification theory, using field data measurements to estimate load models for static and dynamic simulations.

3

Measurement-based Load Modeling Methodology

The compromise between parsimony and flexibility is at the heart of the identification problem

Lennart Ljung

A typical motivation for load modeling estimation is when a bulk power system is connected to distribution system(s) owned by different companies. These companies hardly ever are able to get detailed information of each other's network topologies, but this can be solved with a measurement device installed at the point of common coupling between these different systems. According to the data storage capabilities and the sampling rate of this device, it is possible to approximate the complex smaller systems behavior into equivalent systems, whose parameters can be estimated periodically (and even continuously updated) with the data recorded in the frontier of these two systems.

The proposed load modeling methodology in this work is an application of the measurement-based load modeling, inspired by the system identification procedure. This chapter aims to describe the main tasks involved in obtaining mathematical descriptions from radial subsystems connected to larger systems. The following chapters (4 and 5) will report successful applications of the proposed methodology.

Section 3.1 reviews some of the main concepts that were used to build the framework of the proposed methodology, which is scrutinized in more details in section 3.2.

3.1.

System Identification Approach

It is reasonable to assume that a mathematical model can translate accurately certain features of the real world, whereas it fails to represent other aspects that, by

their turn, may not be important to the desired analysis. The modeling goal, or at least the purpose of it, should be “usefulness”, instead of “truth” [21].

The measurement-based load modeling is essentially a system identification task that builds mathematical models based on real input/output data sets from a system of interest. In this context, a system can be defined as an object from external *stimuli* (input signals or disturbance) that produces observable responses (output signals). Mathematical structures such as differential equations, difference equations, state space models, each with their parameter sets, are models that describe causal relationships between signal inputs and outputs.

3.1.1. System Identification concepts

The following steps summarize the system identification procedure for load modeling:

- The selection of representative input-output system data
- Data treatment for estimation
- The choice of an appropriate model structure
- The estimation of the load model parameters
- The validation of the load model

The terminology used to describe process modeling is not completely standardized, and therefore it is necessary to define some concepts that are going to be useful throughout this thesis. To aid the properly description of the core of this methodology, it will be reviewed some terminology in Table 1 to formulate the concepts in a more rigorous way.

Table 1 – System Identification core concepts [22]

Concept	Symbol	Details
Model	m	This is the mathematical relation between observed quantities, which should be able to predict the properties or behaviors of a system, based on past data.
Parameters set	θ	This is a generic vector of variables contained in the formulation of each model m . Therefore, a model realization could be formulated as $m(\theta)$.

Concept	Symbol	Details
Model Set	M	This is the collection of available or typical models m_k used to describe a particular system, or a class of systems (such as Linear Time-Invariant systems, as discussed in 2.2.4).
Complexity	C	This is a way to compare the size of each of the models m contained in M ; the most used indicator of the model's complexity is the total number of the parameters set θ , i.e., the <i>size</i> or the <i>order</i> of the model (as discussed in 2.2.4).
Estimation Data	Z_e^N	The aim is to find a parameter set θ , for a model m , based on the estimation data set Z_e of size N , which minimizes the differences between the simulated and the measured outputs.
Validation Data	Z_v	This is an approach to ensure the generalization of the model description: the estimated model, “trained” by Z_e , fits well when simulated with validation data Z_v .
Model Fit	$F_{obj}(m(\theta), Z)$	This is the objective function that is formulated to quantify how close the outputs predicted by the estimated model are from the outputs observed in the past. It is usually a scalar value that evaluates error indexes, such as the Mean Squared Error (MSE), or the Mean Absolute Error (MAE).

Fig. 9 illustrates some of the concepts listed in Table 1. In the schematic diagram (a), it is shown that voltage data is the same input for either active power load model or reactive power load model. The models m_p and m_q could have the same mathematical structure (for instance, both ZIP load models), but different parameters sets θ_p and θ_q . The diagram (b) illustrates the concept of a model set M that contains the different models for active power: m_{p1} , m_{p2} and m_{p3} . One of these models could be considered better than the others in a nonparametric approach (one of the possible model selection options in a system identification problem), or they could be tested through a parameter estimation procedure, and the choice for the best load representation could be the best model fit, regarding Table 1, or the best model fit using the validation data.

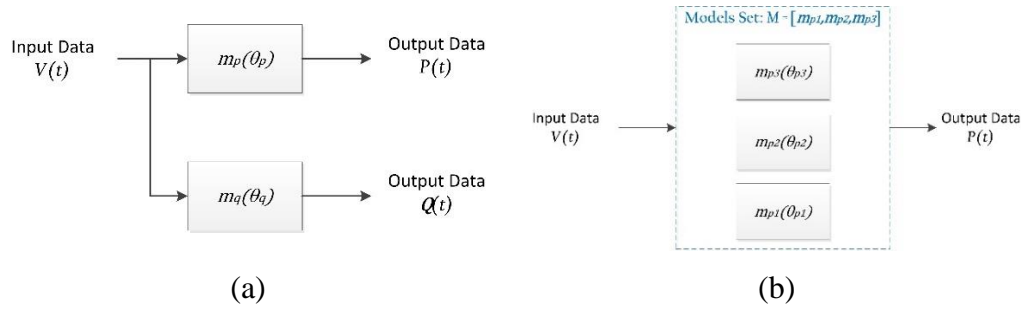


Fig. 9 – Schematic diagrams of the system identification concepts: the diagram (a) shows the concept of voltage data as the input of models m_p and m_q , parameterized respectively by θ_p and θ_q , whose outputs are respectively the active power $P(t)$ and the reactive power $Q(t)$. The diagram (b) represents a models set M containing three different models (and their respective parameters sets) to represent the active power $P(t)$ as a function of voltage input $V(t)$.

3.2. Proposed Load Modeling Implementation

This section describes the proposed load modeling methodology. Since it is inspired by the measurement-based approach, it should be pointed out that this methodology depends on data measurements from a system downstream the point of data acquisition. Fig. 10 illustrates the placement of a disturbance recorder (DR), installed at low side of a transformer at the point of common coupling (PCC) between transmission (upstream system) and distribution systems (downstream, usually strictly radial, system).

It is mandatory to separate the set of data useful to estimate equivalent models of downstream systems. From now on, it is going to be defined the concept of event such as the short-circuits, scheduled voltage control actions or other types of state system changes that lead to load deviations from stable points of operation in the monitored bus bar. The event also has to be originated upstream the DR placement, i.e., in the bulk power system. The reason is that a disturbance originated downstream would change the network topology of the system of interest.

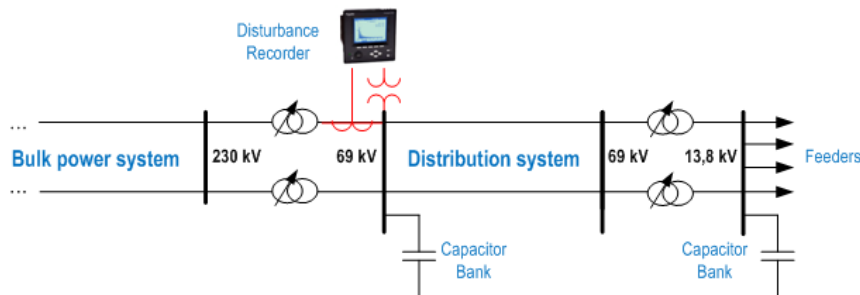


Fig. 10 - Typical placement of disturbance recorders for distribution load model

Other criteria that are common to both case studies, in order to select the useful and representative active and reactive power curves to estimate the load model best parameters, are listed below:

- The duration of the voltage deviation must be large enough, according to the time frame of the analysis.
- The amount of voltage deviation must be “significant”, which means it must be large enough to cause a load response that is also “significant”.
- Pre-disturbance samples should be used for load model’s initialization (P_0 , Q_0 , V_0).
- The downstream system to be modeled must be radial.
- The data must not have any interruption. Missing values should be treated properly.

The diagram of Fig. 11, extracted from [23], represents the framework used in Chapter 4 case study, and the general aspects of it can also describe most of the steps used in the case study of Chapter 5. The diagram presents Power Quality (PQ) data measurements of voltage, active and reactive power that were recorded by Intelligent Electronic Devices (IED) and stored into a PQ database [24]. Then, the data sets are separated into “training data” (N events), for parameter estimation and “validation data” ($Total_Selected_Records - N$ events), to assess the extent of modeling quality, i. e. if the parameters set obtained after fitting the model to the estimation data is able to generalize the validation data.

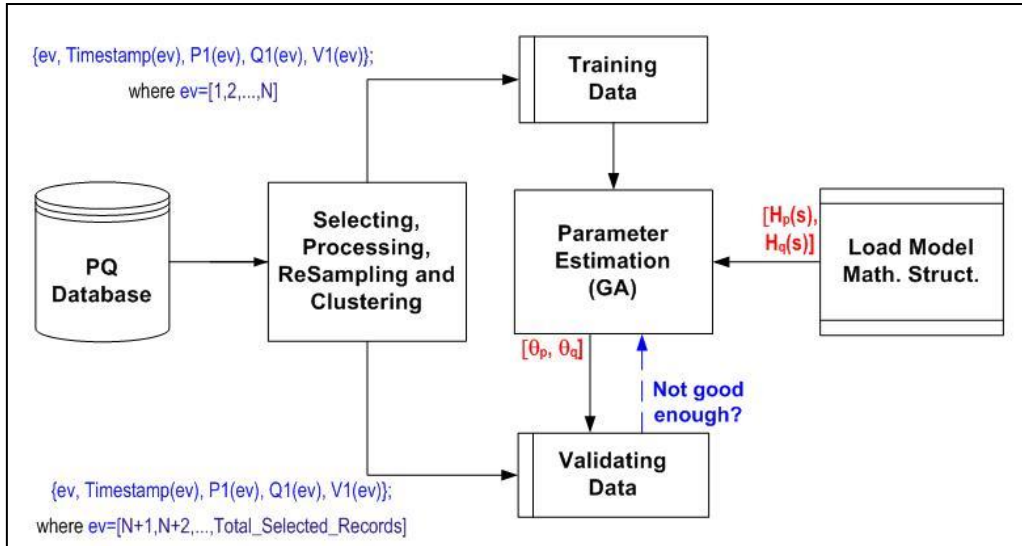


Fig. 11 - Load modeling schematic procedure

The load model structure was selected *a priori* in the case studies described in Chapters 4 and 5. There are many possible reasons for the decision to not test exhaustively all the reviewed load models of chapter 2. In [21], regarding model structure selection: “The choice must be based both on an understanding of the identification procedure and on insights and knowledge about the system to be identified.” which implicitly grants the analyst freedom to determine the criteria for the selection procedure. The type of model set (linear/nonlinear, physical/black-box/neural networks...) and the size can be tested through an iterative process, or the choice simply may be for the availability of the model structure in the target software simulation. In [26], the authors emphasize that it should be employed “the smallest possible number of parameters” (parsimony principle).

In chapter 4, a black-box model was chosen because of its simplicity (small complexity, as defined in Table 1) and also due to the control theory background that helped defining the search space domain and its constraints. In chapter 5, the exponential recovery model was chosen because the theoretical response to input step yields a transient and a steady state output that is suitable to the purpose of the case study.

The parameter estimation methods can be analytical or based on the optimization of an objective function. Algorithms like nonlinear least squares, gradient-based, vector fitting and maximum likelihood can be employed, as well as artificial intelligence techniques such as neural networks, fuzzy logic, simulated annealing and genetic algorithms [27]-[36].

It worthwhile to state that, during the parameter's estimation procedure, more than one technique can be used to search for the best model fit. The Matlab's Simulink Optimization Design is a toolbox that aids such a task, if one implements the selected model structure into this simulation application.

3.2.1. Data treatment

Data treatment involves the observation and selection of system input/output signals, the post-processing, when needed, and if the number of events are large enough, the separation between training and validation data. Filtering of data, removing outliers and the choice of data segments (total samples of an event) should be considered in this procedure step.

In chapter 4, the data consist in instantaneous samples of three phase voltages and currents, recorded at sampling rate of 1920 Hz, before, during and after large voltage variations, i.e., whenever RMS voltage dropped below 0.9 p.u.. These data are processed by a Discrete Fourier Transform (DFT) to obtain voltage and current phasors at fundamental frequency (thus filtering harmonic components), allowing the evaluation of active and reactive power RMS values. In chapter 5, the data consist in RMS voltage, active and reactive power recorded at sampling rate of 1 Hz during one year, and the events selection criteria separates, through an algorithm *ad-hoc*, the voltage steps performed by transformers with tap changers.

3.2.2. Model selection

A load model structure m_k (Table 1) is a mathematical mapping of input/output data, whether it is a SISO model or a Multi-Input Multi-Output (MIMO). One of the key issues on system identification task is whether to choose an appropriate model structure *a priori* or direct estimate of the model by computing the impulse response of the system or by correlation analysis [26]. The first one is known as *parametric method* and consists on specifying a mathematical model that can be parameterized to fit a set of training data. The second one is known as *nonparametric method*, and do not assume or impose any model structure [21]. Fourier analysis, correlation analysis and frequency-response analysis are examples

of nonparametric methods for model selection. This work adopted the *parametric method*.

Model selection can be also the process of choosing a mathematical description m_k of a phenomenon among a set of model structures M , as defined in Table 1. This choice can be oriented by some heuristic, previous knowledge from specialists, other known physical models, or even by a mixture of different components. After the specification of a model structure, it should be estimated the parameters set θ that best approximate the system output to the data output observed. This is achieved by minimizing the mismatch between estimation data and the simulated data. It is also desirable to obtain a model among a set of models that is not too complex. The equation (3.1) illustrates that a good model estimative m_k should find m within M as “simple” and accurate as possible, which means that small order models are preferable than the more complex ones.

$$m_k = \arg \min [F_{obj}(m_k(\theta), Z_e^N) + f(C(m_k), N)] \quad (3.1)$$

where: $m_k \in M$

3.2.3. Parameter Estimation

The technique chosen to estimate load model parameters was Genetic Algorithm (GA), which is a heuristic-based method inspired on natural selection and biological evolution, described by Charles Darwin in his Evolutionary Theory. The common underlying idea behind GA is: a random set of feasible vector-solutions is generated and structured in real-valued vectors. The whole set of these structured vectors is called *population* and each vector is called an *individual*, or a *chromosome*. Individuals that are well evaluated by a *fitness function* or the *objective function* to be minimized, formulated in (3.2), are more likely to persist between algorithm iterations (*generations*’ evolution).

Fitness evaluation of each individual allows a way to identify the best solutions. The individuals that give best solutions receive higher scores and are most likely to “survive” through generations (iterative process), because the best individuals have a higher probability of being selected and used to form next generation of population. A part of the individuals within a population, in a given generation, is recombined (*crossover*), and another part is modified in one of its

elements (*mutation*). Every generation is created through selecting, recombining and mutating individuals from the previous generation.

Through iterative process, the fitness of the best chromosome improves as well as the total fitness of the population as a whole. The population "evolves" toward a quasi-optimal solution, over a finite number of generations or until a convergence tolerance has been reached. GA can be applied to solve a variety of optimization problems that are not well suited for standard optimization algorithms, including those in which the objective function is discontinuous, non-differentiable, stochastic or highly nonlinear [37].

The objective function to be minimized is the relative mean square error between measurements and the output of load models, as shown in (3.2)).

$$\begin{aligned} F_{obj}(\theta_p) &= \frac{1}{N} \sum_{ev=1}^N \frac{1}{n} \sum_{t=1}^n \left(\frac{P_{meas}(ev, t) - P_{calc}(ev, t, \theta)}{P_{meas}(ev, t, \theta)} \right)^2 \\ F_{obj}(\theta_q) &= \frac{1}{N} \sum_{ev=1}^N \frac{1}{n} \sum_{t=1}^n \left(\frac{Q_{meas}(ev, t) - Q_{calc}(ev, t, \theta)}{Q_{meas}(ev, t, \theta)} \right)^2 \end{aligned} \quad (3.2)$$

Where:

- F_{obj} – the objective function to be minimized.
- θ_p and θ_q are active and reactive load model parameters' vectors, respectively.
- N – the total of events recorded, typically an event is any "significant" load input deviation (the "significant" depends on the load model purpose, because the expected).
- ev – the variable that indexes each of the N events.
- n – the total of samples recorded during an event.
- t – the variable that indexes the order of the samples.
- $P_{meas}(ev, t)$, $Q_{meas}(ev, t)$ – active and reactive power data measured for the event ev , with n samples.
- $P_{calc}(ev, t, \theta_p)$, $Q_{calc}(ev, t, \theta_q)$ – active and reactive power data calculated by any load model structure for the event ev , with n samples, and parameterized by θ_p and θ_q .

3.2.4. Model Validation

After parameter estimation using the training data, the model validation consists in the assessment of model quality, using a different data set (validation data set). This is achieved simulating the validation data inputs into the estimated load model, and comparing the simulated outputs with the measured outputs from the validation data set, whether visually by plotting simulated and measured outputs, or by comparing objective functions values.

It is expected that the overall error from validation data simulation will be greater than the training data, but besides the verification if the model fits well observed model outputs, the real goal is to achieve a good enough model for the intended purpose, which means that the accuracy depends on the model usage and the importance of it within the overall analysis of simulation.

A more rigorous analysis test, from a statistical point of view, relies on the residual's analysis, in order to relate residuals to random error, which essentially implies that the model fit is well explained by predictor variables and the nature of unpredictability is due to noise, or any other inherent real-world randomness. Residual plots for good model fits presents zero mean data and small variance.

3.3. Summary of the chapter

This chapter presented an overview of the load modeling proposed methodology, which adopted the measurement-based approach and system identification concepts. Voltage and demand data are essential to build an equivalent representation of complex electrical network topologies into single subsystems connected to a bigger system. The guidelines of the system identification procedure are flexible enough to combine different techniques in each step, and Matlab is very helpful since its toolboxes contain probably most of the most common algorithms for model selection, parameter estimation and graphical analysis.

The concepts presented in this chapter were applied in the case studies described next (chapters 4 and 5); although the purpose of the load modeling in both cases are different, the steps of the system identification procedure are pretty much

the same. The selection of model structures was made *a priori* and the graphical results have shown negligible mismatches between model output observations and simulations.

Therefore, the system identification can be understood as the framework for estimating load models, neither depending on the choice of estimation parameters technique, nor if the model selection followed a parametric or nonparametric investigation. The main steps should be followed, but on each one of them, there are more than one option that can be used. This makes the proposed methodology flexible enough to allow the choice of different parameter estimation techniques for different modeling problems, different model selection approaches, and even different signal processing techniques.

Nevertheless, there is an order to perform the steps required in the search of accurate models, and the answer for this particularly complicated philosophical aspect (what makes a good model? What kind of scenarios this model is suitable, and what are its limitations?), is always associated to the type of simulation that expects to use one or more load models that represent, as close as possible, complex systems into equivalent subsystems.

4

Load Modeling for Dynamic Simulation – A Case Study

This chapter will present a case study of the load modeling methodology proposed in chapter 3 for dynamic simulation. The data used was obtained from measurements due to large voltage variations from a Brazilian transmission utility. The purpose of the study was to perform dynamic simulations, using a Brazilian software, representing distribution systems with measurement-based load models. This software can perform electromechanical transient analysis, using input data from the topology of the interconnected transmission system of Brazil published by the national operator system. The distribution systems are not represented in detail within the input data published by the national operator system, and the measurement-based load modeling methodology proposed in chapter 3 provides a systematic procedure to develop equivalent distribution systems representations.

In this chapter, section 4.1 presents a thorough description of the case study, section 4.2 describes the signal processing techniques involved in filtering voltage and current instantaneous samples into positive sequence components of voltage, active and reactive power, since electromechanical transient analysis assumes the system is balanced. Section 4.3 presents a load modeling structure that has no physical meaning, and thus it is designed to cover a wider range of load response behaviors. Section 4.4 details how the parameter estimation was solved by the algorithm genetic optimization procedure, separating two practical approaches to determine the estimation goals: the first one (subsection 4.4.1) aims to find a parameters set, that is a load model, as accurate as possible to describe load response from a set of disturbances, recorded in different scenarios; the second approach focus in describing a single contingency, in order to reproduce a special event of interest. Finally, section 4.5 shows the transient electromechanical simulation of the effect of a temporary fault in the transmission system, observing the frontier between the transmission system and a distribution system that contains generation within its topology.

4.1. Case Study Description

Fig. 12 presents the grid of the case study described in this chapter. The green lines represent 230 kV transmission system and red lines represent 500 kV. Black triangles represent hydraulic generating units and black dots indicate power quality monitors installed in substations, on both sides of 230-69 kV.

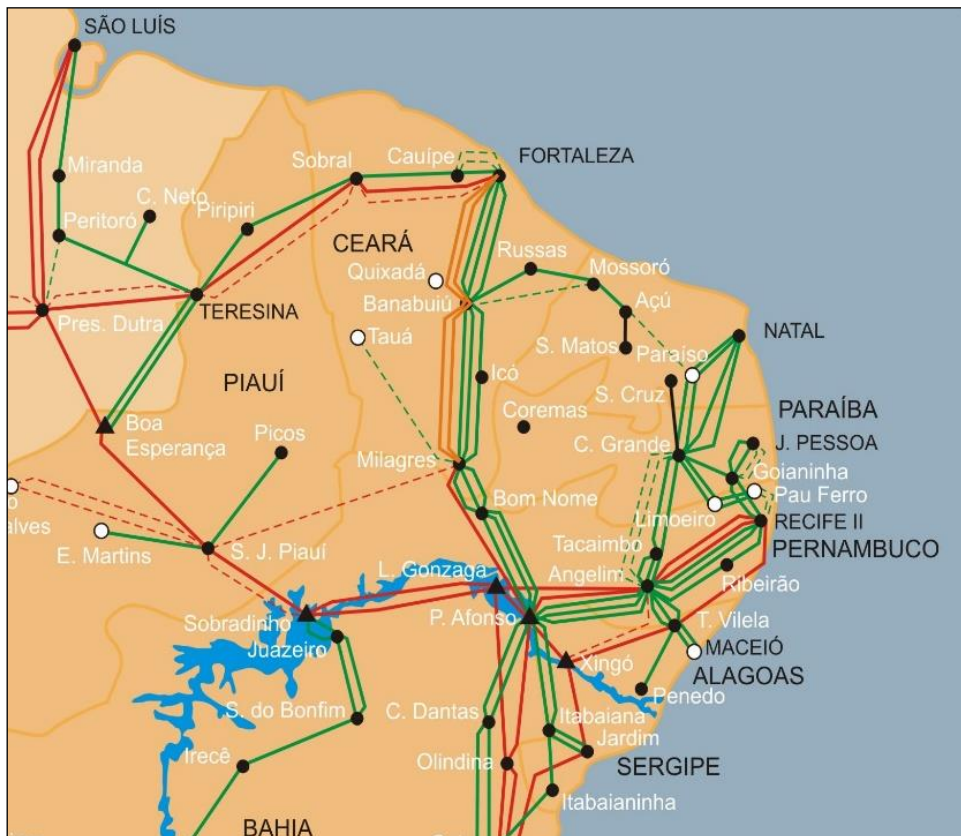


Fig. 12 - A power electric grid diagram from a northeastern Brazilian utility

The low-voltage side of substation 230-69 kV transformers is considered the ideal place for collecting load data [25]. Intelligent Electronic Devices (IED), spread over the transmission grid, send data through an Ethernet network to a centralized server. Using suitable criteria to select typical disturbances (events) that are more likely to occur in a particular load bus, load models that cover the most common scenarios can be estimated to simulate electromechanical transients in a more realistic way.

Using the proposed methodology, presented in chapter 3, many distribution systems were modeled, and some results will be shown in this chapter. However, there is a particular one that provided the most remarkable insights about the load

model structure chosen. Fig. 13 reveals a schematic diagram for this particular distribution system that is going to be used in this case study. It is an Active Distribution Network (ADN), although the generators are thermoelectric generators, different from the example that motivated the load model formulated in section 2.2.3. The results of this system will be presented in section 4.4.

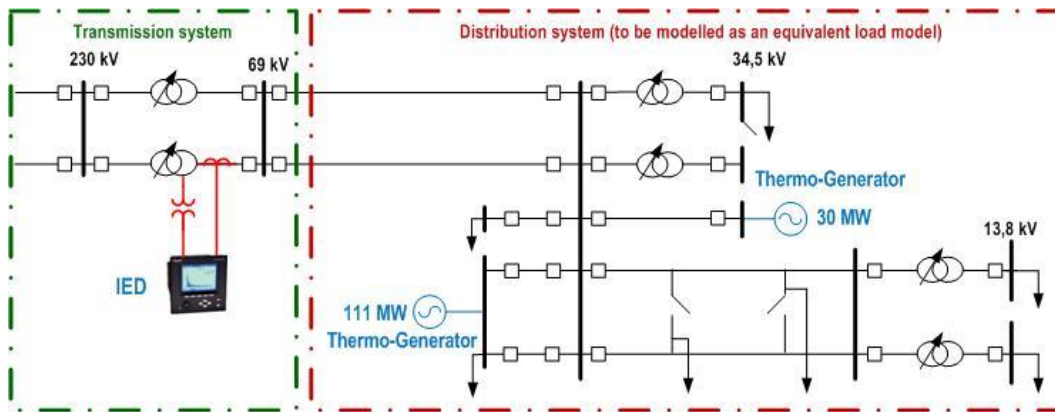


Fig. 13 – Case study measurement setup

4.2. Data treatment

Since our main goal is to explore a causal relationship between power and voltage, it is necessary to perform some signal processing and transformation. The load models must be implemented into electromechanical transient software, and the latter evaluates positive sequence networks, thus it is necessary to perform Fortescue Transformation on current and voltage phasors data sets. Therefore, positive sequence components of active and reactive power are the output data selected for the dynamic load model case study, while positive sequence components of voltage are the input data for the dynamic load model. This section describes this signal processing that transforms instantaneous voltage and current data samples into voltage, active and reactive power positive sequence components.

Instantaneous three phase voltages and currents must be post-processed by means of a Discrete Fourier Transform (DFT) [11], based on the sliding window algorithm, for obtaining complex voltages and currents phasors. So, these signals are transformed into zero, positive and negative sequence components through Fortescue transform. The reason to perform this data post-processing is related to the nature of the electromechanical analysis, which assumes the transmission

systems are balanced, and thus it is modeled as a single phase, or positive sequence network.

Equation (4.1) presents the instantaneous voltages and currents formulated in the continuous time domain. Equation (4.2) formulates the RMS equivalent equivalence of voltages and currents, regarding the continuous-time waveform, using the polar coordinates. DFT algorithm [19] is able to extract all the harmonic contents within a signal, but equation (4.3) evaluates only the fundamental frequency setting $H=1$.

$$v(t) = V_{max} \cos(\omega t + \theta_v) \quad i(t) = I_{max} \cos(\omega t + \theta_i) \quad (4.1)$$

$$V_{RMS} = \frac{V_{max}}{\sqrt{2}} e^{j\theta_v} \quad I_{RMS} = \frac{I_{max}}{\sqrt{2}} e^{j\theta_i} \quad (4.2)$$

$$V_{RMS}(t) = \frac{\sqrt{2}}{N} \sum_{n=t}^{t+N-1} v\left(\frac{n\Delta t}{N}\right) e^{-j\left(\frac{2\pi n H}{N}\right)} \quad I_{RMS}(t) = \frac{\sqrt{2}}{N} \sum_{n=t}^{t+N-1} i\left(\frac{n\Delta t}{N}\right) e^{-j\left(\frac{2\pi n H}{N}\right)} \quad (4.3)$$

Where:

- $v(t)$ and $i(t)$ are the instantaneous voltage and current discrete samples recorded by IED.
- V_{max} and I_{max} are the amplitudes of the sinusoid signals of voltage and current, respectively.
- ω , θ_v and θ_i are respectively angular frequency, in radians per second and the signal phases of voltage and current signals.
- V_{RMS} and I_{RMS} are the root mean square for voltage and current signals.
- N is the number of samples in one fundamental cycle.
- Δt is the sampling period.
- n is the index of the sample within one fundamental cycle.
- H is the harmonic order evaluated by DFT.
- t is the variable that indexes time.

IED records current and voltage waveforms sampled at 1920 Hz, which means each waveform contains 32 samples per cycle (60Hz). Fig. 14 presents an example of instantaneous voltages and currents, recorded at the low side of a 230/69kV transformer.

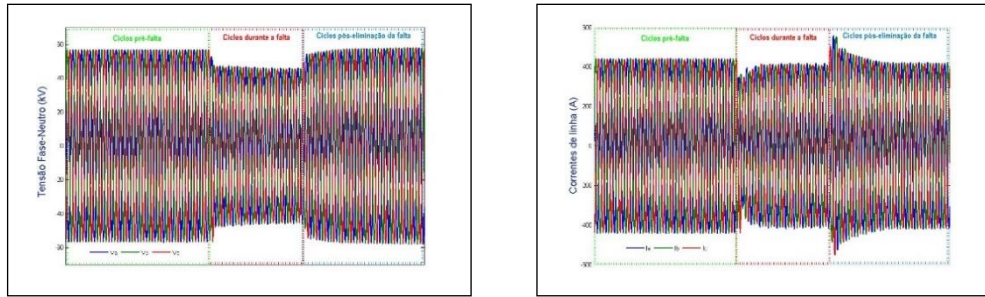


Fig. 14 - Instantaneous voltages and currents recorded by the IED with a sampling rate of 1920 Hz

Fig. 15 presents the results of (4.3) for a recorded contingency, where the left graphic shows positive sequence active power component data (left axis, blue curve), and voltage data (right axis, green curve) associated, whilst right graphic shows reactive power (left axis, red curve) against the same voltage data recorded from a low voltage side of a step-down transformer, as shown in the diagram of Fig. 10.

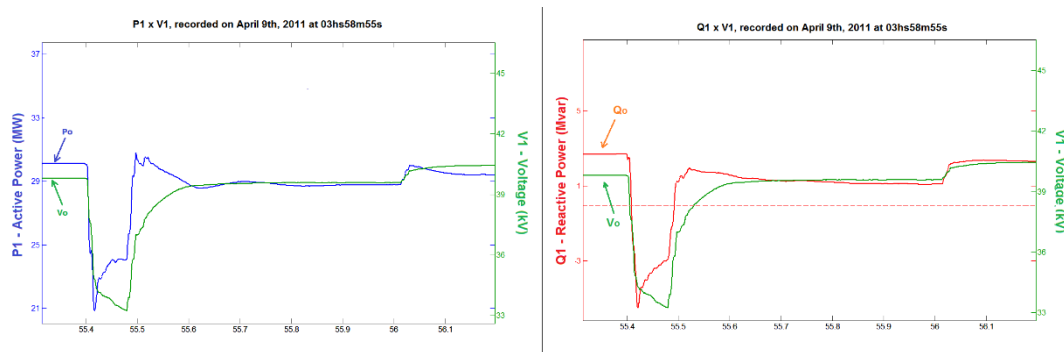


Fig. 15 - Positive sequence active (blue) and reactive (red) power components against positive sequence voltage (green) component.

Fig. 16 shows only the result of DFT onto one cycle in polar coordinates, although the algorithm is performed through the whole waveform using a sliding window. DFT may obtain magnitude of different harmonic orders, but this work deals only with fundamental frequency, so DFT can be thought as a filter of all harmonic components other than the fundamental frequency (60Hz).



Fig. 16 - Discrete Fourier results from snapshots of a pre-disturbance (green phasors) and “during the disturbance” (red phasors) voltages and currents shown in Fig. 15.

The phasor form in (4.3) can be decomposed in rectangular coordinates, evaluating complex numbers that can be used to calculate the component sequence phasors, by means of the Fortescue transform. Finally, equation (4.5) formulates the three-phase active and reactive power components and (4.6) computes the positive sequence components.

$$\begin{bmatrix} V_0 \\ V_1 \\ V_2 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & a & a^2 \\ 1 & a^2 & a \end{bmatrix} \begin{bmatrix} V_a \\ V_b \\ V_c \end{bmatrix} \quad \begin{bmatrix} I_0 \\ I_1 \\ I_2 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & a & a^2 \\ 1 & a^2 & a \end{bmatrix} \begin{bmatrix} I_a \\ I_b \\ I_c \end{bmatrix} \quad (4.4)$$

$$S = P + jQ = VI^* = V_a I_a^* + V_b I_b^* + V_c I_c^* \quad (4.5)$$

$$P_1 = 3 \frac{V_1}{\sqrt{2}} \frac{I_1}{\sqrt{2}} \cos(\delta_{V_1} - \delta_{I_1}) \quad Q_1 = 3 \frac{V_1}{\sqrt{2}} \frac{I_1}{\sqrt{2}} \sin(\delta_{V_1} - \delta_{I_1}) \quad (4.6)$$

Where:

- V_0 and I_0 are zero sequence voltage and current values, respectively.
- V_1 and I_1 are positive sequence voltage and current values, respectively.
- V_2 and I_2 are negative sequence voltage and current values, respectively.
- V_a , V_b and V_c are the three line to neutral voltages.
- I_a , I_b and I_c are the three line currents.
- a is an operator that shifts phase of the complex numbers by 120° .
- S , P and Q are respectively apparent, active and reactive powers.
- P_1 and Q_1 are the positive sequence active and reactive powers.

Fig. 17 summarizes the data treatment.

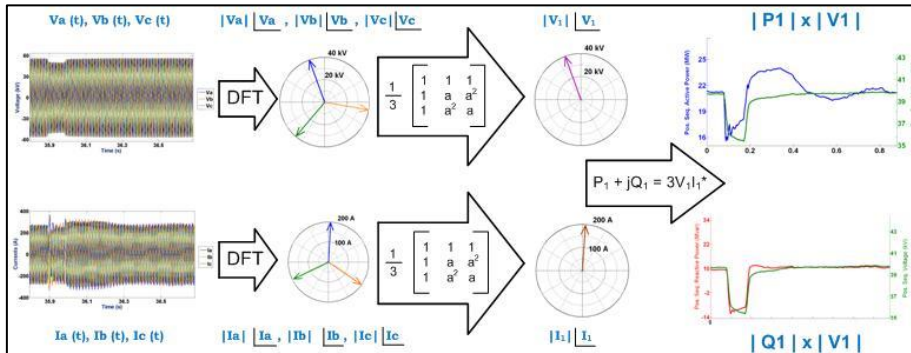


Fig. 17 - Processing data procedure

4.3. Model Selection

The model selection for this case study was guided by the following reasons:

- It was mandatory to choose a model able to be implemented in a simulation software.
- It was desirable to choose a simple model, but capable of reproducing the load response phenomena accurately enough for dynamic simulations purposes.
- It was mandatory to understand the parameters' scope, which means that the parameters theoretical constraints led to restrict the solution search space, aiding the parameters search algorithm to be efficiently designed.
- It was also desirable to choose a load model classified as dynamic, rather than a static one, since the main purpose was to use it for dynamic simulations
- There was no need to choose a load model with physical meaning, and a purely mathematical one was designated targeting at reproducing data measurements, covering as many different scenarios as possible

Recalling the black box load model presented in (2.43), if frequency and noise data are not included in that formulation and also making $np=nv=nq=2$, then the resulting second-order load model is (4.7):

$$\begin{aligned} P(t) &= P_0 \left[\sum_{m=1}^2 \omega_{mp} \frac{P(t-m)}{P_0} + \sum_{m=0}^2 \omega_{mvp} \frac{V(t-m)}{V_0} \right] \\ Q(t) &= Q_0 \left[\sum_{m=1}^2 \omega_{mq} \frac{Q(t-m)}{Q_0} + \sum_{m=0}^2 \omega_{mvq} \frac{V(t-m)}{V_0} \right] \end{aligned} \quad (4.7)$$

Where:

- t is the integer variable that indexes data sampled at equal discrete intervals.
- m is the integer that indicates the number of past samples that are taken into account to evaluate the actual power.
- V_0 , P_0 , Q_0 are respectively voltage (in volts), active (in watts) and reactive power (in vars) samples in pre-fault state.

- $V(t)$, $P(t)$, $Q(t)$ are respectively voltage (in volts), active (in watts) and reactive power (in vars) at instant t .
- ω_{1p} , ω_{2p} , ω_{0vp} , ω_{1vp} , ω_{2vp} , ω_{1q} , ω_{2q} , ω_{0vq} , ω_{1vq} , ω_{2vq} are the parameters to be estimated by the optimization technique.

It was possible to reduce parameters set, since the steady state conditions correspond to the initial conditions $P(t=0)=P_0$ and $Q(t=0)=Q_0$, which means that, for each model, the sum of parameters equals one. So,

$$\begin{aligned}\omega_{2vp} &= 1 - \omega_{1p} - \omega_{2p} - \omega_{0vp} - \omega_{1vp} \\ \omega_{2vq} &= 1 - \omega_{1q} - \omega_{2q} - \omega_{0vq} - \omega_{1vq}\end{aligned}\quad (4.8)$$

Fig. 18, extracted from [26], associate the parameters` domain with the stable poles of the discrete time transfer function (roots of the denominators), and also shows the respective autocorrelation and partial autocorrelation functions. Observe that the parameter ϕ_1 in the graphic is ω_{1p} and ϕ_2 is ω_{2p} in (4.7).

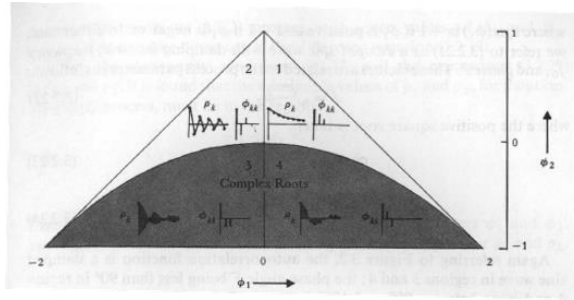


Fig. 18 – Parameters $\omega_{2p,q}$, $\omega_{2p,q}$ boundaries (ϕ_1 and ϕ_2 respectively) [26]

Parameters are subjected to the constraints of (4.9), in order to guarantee the stability of the model.

$$\theta_p = [\omega_{1p}, \omega_{2p}, \omega_{0vp}, \omega_{1vp},]$$

Subject to: $-1 < \omega_{2p} < 1$

$$\omega_{2p} - \omega_{1p} < 1$$

$$\omega_{2p} + \omega_{1p} < 1$$

$$\theta_q = [\omega_{1q}, \omega_{2q}, \omega_{0vq}, \omega_{1vq},] \quad (4.9)$$

Subject to: $-1 < \omega_{2q} < 1$

$$\omega_{2q} - \omega_{1q} < 1$$

$$\omega_{2q} + \omega_{1q} < 1$$

The load model presented in (4.7) is the discrete time-domain counterpart of ordinary differential equations. This formulation is a set of difference equations, and the system outputs active and reactive powers $P(t)$ and $Q(t)$ are evaluated taking into account past samples of the system respective outputs, as well as the actual and past samples of voltage samples $V(t)$ inputs. These past samples are precisely what characterizes this load model as a “dynamic” one, and there are a lot of literature developments regarding this area.

The next subsection presents some of the knowledge about transfer functions, and how it is interesting to formulate this model as a discrete time transfer function or as a continuous time transfer function, using suitable mathematical transformations, such as Z transform, Laplace Transform and Tustin transform.

4.3.1. Transfer Functions

Transfer functions are rational polynomial functions that relate input(s) and output(s) of dynamical systems in a compact way. The ratio of polynomials can be differential equations (continuous time domain) or difference equations (discrete time domain). For systems of finite dimension, the transfer function may be defined as a rational function of a complex variable.

“A Linear Time-Invariant (LTI) system with a rational system function has the property that the input and output sequences satisfy a linear constant-coefficient difference equation. Since the system function is the Z transform of the impulse response, and the difference equation, satisfied by the input and output, can be determined by inspection of the system function, it follows that the difference equation, the impulse response, and the system functions are equivalent characterizations of the input-output relation of a LTI discrete-time system” [19].

The Z transform plays the same role in discrete time domain as the Laplace Transform for continuous time domain. If the Z transform is applied to equation (4.7), the load model assumes a transfer function form in discrete time domain, as it is shown in the first column of

Table 2. The continuous time domain counterparts is also presented in the second column.

pasta samples of the input (voltage signal). The Laplace transform is similar to the Fourier transform. While the Fourier transform of a function is a complex function of a real variable (frequency), the Laplace transform of a function is a complex function of a complex variable.

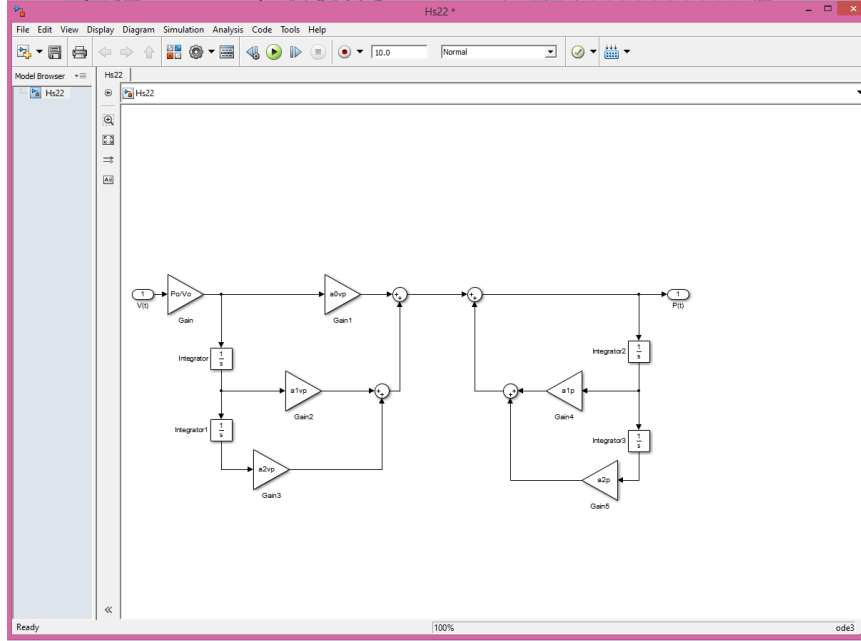


Fig. 20 - Simulink implementation of the continuous time transfer function of the proposed active power load model

In continuous time domain, an equivalent formulation of the transfer function presented in

Table 2 is presented in (4.10):

$$P(s) = \frac{P_0 \prod_{sp=1}^{nv} (s - z_{sp})}{V_0 \prod_{sp=1}^{np} (s - p_{sp})} V(s) \quad Q(s) = \frac{Q_0 \prod_{sq=1}^{nv} (s - z_{sq})}{V_0 \prod_{sq=1}^{nq} (s - p_{sq})} V(s) \quad (4.10)$$

Where:

- s : Laplace complex frequency
- $s = \sigma + j\omega$: Laplace complex frequency, where the real part σ is the decaying component of impulse response and ω is the oscillatory component of the same impulse response.
- $P(s)$, $Q(s)$ and $V(s)$: Laplace transform of the active power, reactive power and voltage, respectively;
- P_0 , Q_0 and V_0 : steady state values of active and reactive powers and voltage, respectively;

- z_{sp}, z_{sq} : ***nv*** zeros for each active and reactive transfer functions, corresponding to the frequencies for which the value of the transfer function's numerator becomes zero.
- p_{sp}, p_{sq} : ***np*** (or ***nq***) poles for active (or reactive) transfer functions, corresponding to the frequencies for which the value of the transfer function's denominator becomes zero.

Poles and zeros are constrained to a certain range of acceptable values to ensure transfer function stability [19], [21] and [26] and as stated in

Table 2. Poles effectively define a differential system with homogeneous response. Thus, the unforced response of a linear SISO, subject to the initial conditions (P_0, Q_0) , the homogeneous response is a summation of exponential functions whose decaying ratios are the poles of the characteristic equation, that is the transfer function's denominator roots. Next, it will be listed all the possibilities, including unstable poles, and the corresponding differential system unforced response.

When the poles are:

- **Complex with real negative parts** in the transfer function, defined in the frequency domain, the corresponding impulse response in time domain is a sine with decaying amplitude accordingly (*underdamped system*).
- **Complex with real positive parts** in the transfer function, defined in the frequency domain, the system outputs an oscillatory impulse response with increasing amplitude (*unstable system*).
- **Complex with real parts equal to zero**, the transfer function impulse response is a sine with constant amplitude (*undamped system*).
- **Real negative**, transfer function's impulse response is a decaying exponential (*overdamped system*, if they are different, and *critically damped*, if they are equal).
- **Real positive**, transfer function's impulse response is an increasing exponential (*unstable system*).

These effects are illustrated in the Fig. 21:

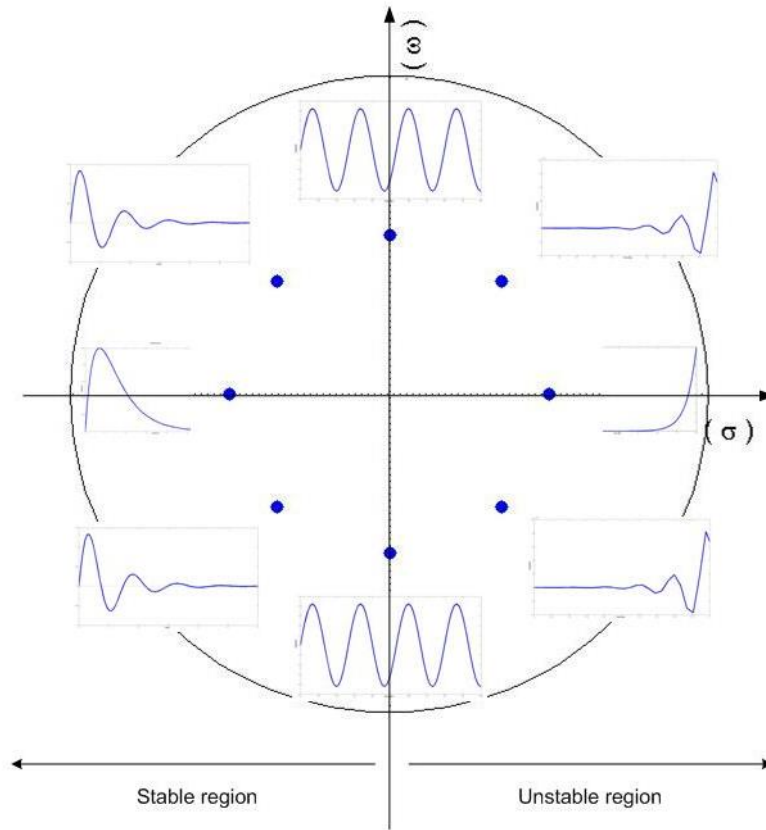


Fig. 21 - Analysis of homogeneous response from the system pole locations.

It is also analyzed that poles in the left plane with big magnitude are the ones decays faster. Regarding the s-plane in Fig. 21, poles far from the origin in the left-half plane correspond to components that decay rapidly, while poles near the origin correspond to slowly decaying components (dominant poles or dominant long term response components in the overall homogeneous response).

In (4.11) there are presented two sets of parameters for active and reactive load models formulated for the higher order transfer functions presented in (4.10). It is worth to highlight that the number of parameters depends on the orders np and nq of the respective load models. No matter if the transfer function's order is odd or even, it is trivial to realize that the solution of the characteristic equation yields complex poles/zeros, they will always appear as complex conjugated.

$$\begin{aligned} \theta_p(np) &= [z_{1p}, z_{2p}, \dots, z_{np}, p_{1p}, p_{2p}, \dots, p_{np}] \\ \theta_q(nq) &= [z_{1q}, z_{2q}, \dots, z_{nq}, p_{1q}, p_{2q}, \dots, p_{nq}] \end{aligned} \quad (4.11)$$

4.4. Parameter Estimation

- The load model (4.7) is considered in [21] as “perhaps the most basic relationship between the input and output” and it is classified in the same reference as the Auto-Regressive with an eXogeneous input (ARX). The boundary limits of its parameters set were registered in [26], as it was shown in Fig. 18 and in (4.9). In order to pursue a simple model, it was the first choice for the parameter estimation procedure and its results were able to cover many different load response scenarios, in different measurement sites.

The second order continuous transfer function formulated in

- Table 2 was also used in the parameter estimation procedure. Although the available data was sampled from the Analog/Discrete (A/D) system of each IED, and therefore all available data are discrete, throughout the parameter estimation procedure it was always necessary to convert the continuous transfer function back to the discrete transfer function, in order to calculate the error between the model output and the measured active/reactive power data.

The parameter estimation for this case study has explored more than one of the options presented in section 4.3, both listed above. There are two basic approaches regarding either discrete transfer function, or continuous transfer functions, whereas each of which also may be formulated in two different ways. Table 3 summarizes the formulations of section 4.3:

Table 3 – Summary of discrete and continuous time transfer functions to represent aggregated load

Discrete-Time Domain Transfer Function	Continuous-Time Domain Transfer Function
$P(z) = \frac{P_0 \prod_{zm=1}^2 (z + z_{zm})}{V_0 \prod_{zm=1}^2 (z + p_{zm})} V(s)$	$P(s) = \frac{P_0 \prod_{sm=1}^2 (s + z_{sm})}{V_0 \prod_{sm=1}^2 (s + p_{sm})} V(s)$
Subjected to: $ p_{zm} < 1 \quad zm=1,2$	Subjected to: $\text{RE}\{p_{sm}\} < 0 \quad sm=1,2$
$P(z) = \frac{P_0}{V_0} \left[\frac{\omega_{0vp} z^2 + \omega_{1vp} z + \omega_{2vp}}{z^2 - \omega_{1p} z - \omega_{2p}} \right] V(z)$	$P(s) = \frac{P_0}{V_0} \left[\frac{\alpha_{0vp} s^2 + \alpha_{1vp} s + \alpha_{2vp}}{s^2 + \alpha_{1p} s + \alpha_{2p}} \right] V(s)$
$\begin{aligned} -1 &< \omega_{2p} < 1 \\ \omega_{2p} - \omega_{1p} &< 1 \\ \omega_{2p} + \omega_{1p} &< 1 \end{aligned}$	$\alpha_{2vp} = \alpha_{2p}$

There are two main approaches for choosing the training data to estimate a parameter set. One may need a parameter set that best describes the power behavior for the contingencies that are most likely to occur (in this context, these were considered the *typical contingencies*). In other words, this approach intends to find, within a given parameter's solution search space, a load model that is supposed to simulate the power dynamics for a set of typical contingencies recorded by a monitoring system. This is examined in section 4.4.3.

Finding a load model for typical contingencies, in the context explained above, means to estimate a parameter set as accurately as possible to simulate different voltage sags, which means different types of voltage sags (one-, two- or three-phase drops), different severities (minimum voltage during disturbance) and different scenarios (different pre-disturbance load levels P_0 and Q_0 , different hour of the day, day of week, etc.). Therefore, the choice of training data set and validation data set for estimating an equivalent load representation is crucial for obtaining a general load model that is expected to cover the most typical scenarios.

The other approach is to reproduce one event by using data recorded before, during and after this specific event of interest, which means, in this case, that $N=1$ for the fitness function formulated in (3.2).

It is sometimes necessary to better understand a specific contingency in a post-mortem analysis. For this case, whenever a record of this specific disturbance is available, it is considered the best way to estimate an equivalent load model to represent that subset of the main system. If it is not available, it should be tried to use another event as similar as possible to the one intended to be modeled. This is examined in section 4.4.4.

To find a parameter set that best simulates different contingencies, one must retrieve from the database some representative records to act as training data. Their indicators (minimum RMS voltage, duration, type...) must range within an expected value, characterizing a typical behavior of a specific distribution system to be modeled.

Next, it will be presented an example, showing the evolution of the parameter estimation process. Fig. 22, Fig. 23 and Fig. 24 present the evolution of the optimization process implemented through GA for the estimation of the parameters set, subjected to the constraints of (4.9). It was selected 4 contingencies (*events*), recorded after the voltage has dropped below 0.9 p.u. during a few cycles and it can

be seen the green curves that show the positive sequence components of voltage variations recorded, the blue curves that show the positive sequence components of the active power load response and the red curves that show the active power, simulated by an adaptation of the load model formulated in (4.7). At the top of each figure, the load model equations show the best estimates of the GA population generations. At the left bottom of each figure, the objective/fitness function's evaluations, as formulated in (3.2) and the corresponding parameter set, as formulated in (4.9).

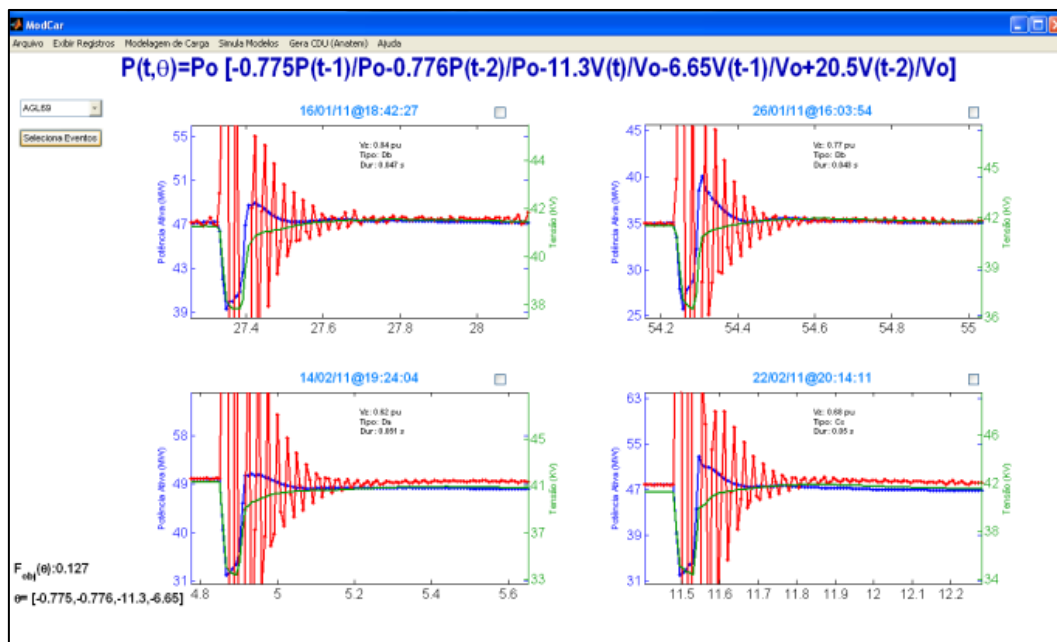


Fig. 22 – Estimating parameters of active power (first generation)

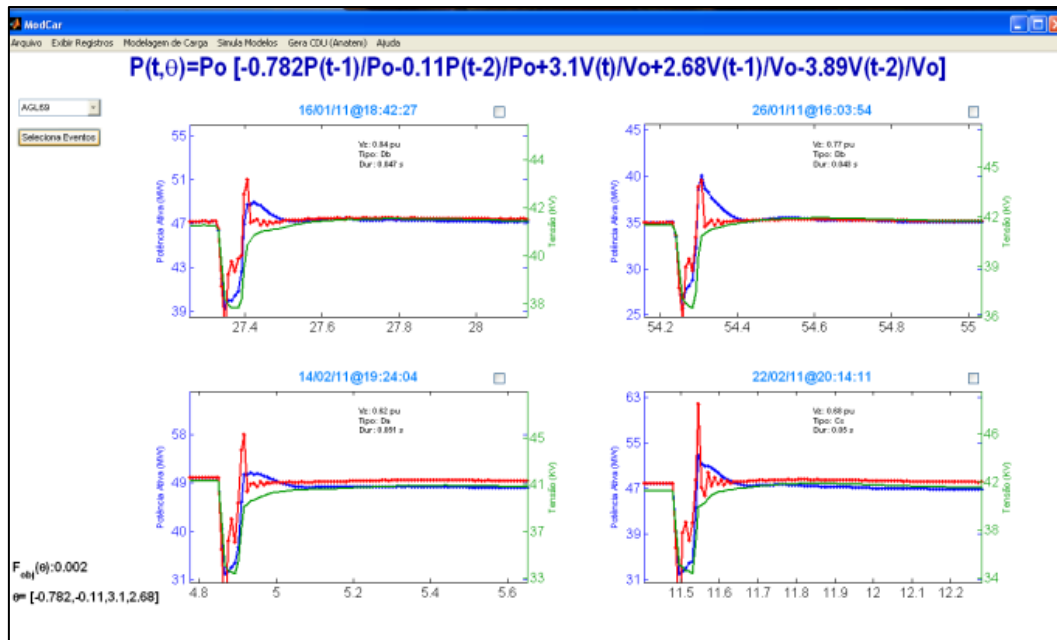


Fig. 23 - Estimating parameters of active power (fourth generation)

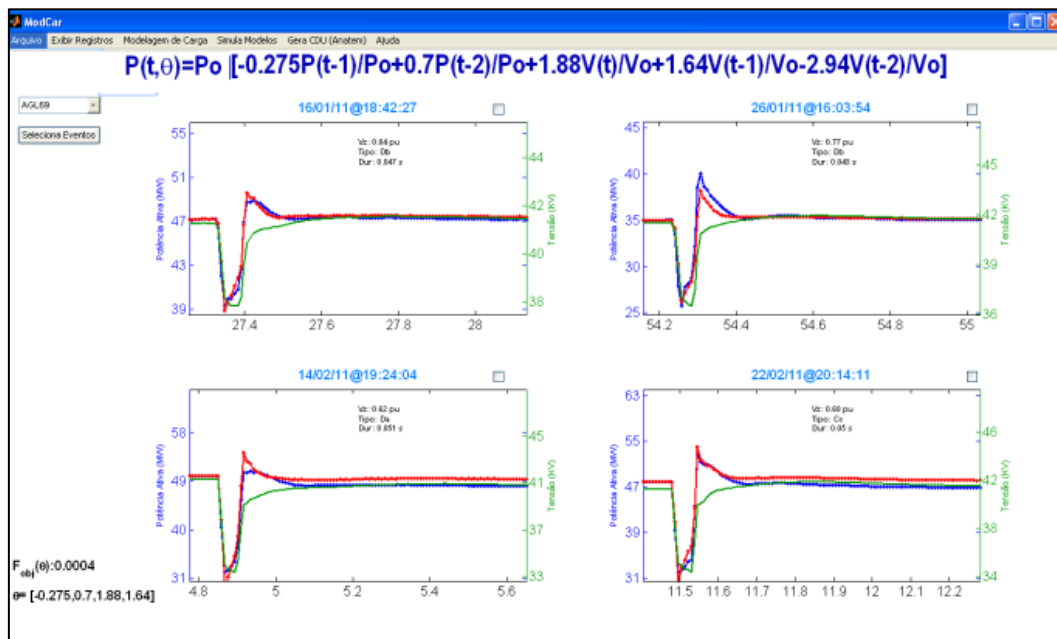


Fig. 24 - Estimating parameters of active power (after 200 generations)

4.4.1. Genetic Algorithm

Fig. 25 shows a flexible coding of the solution for the optimization problem, using the parameters set of (4.11). Genetic Algorithm (GA) is the optimization procedure that will search for parameters that best suits the training data. GA formulation uses concepts of the theory of evolution from Charles Darwin, naming

individual or *chromosome* a vector containing a solution. In this work, there were two different codification solutions to set GA for the parameter estimation. The first *chromosome* is a straightforward vector, which means that every parameter is stored in an element (*chromosome genes*) of the solution vector (*chromosome/individual*), within the *population of solutions* or *individuals*. The other one, contains two extra elements, that informs how many real zeros (nZ) and how many real poles (nP) the transfer function contains; the remaining elements are the parameters. The algorithm implementation can interpret then, what are the zeros and poles, for the fitness function score that solution.

Chromosome for the 2nd order model:

Upper bounds	2	2	0	0	0	0
	nZ	nP	z1	z2	p1	p2
Lower bounds	0	0	$-\infty$	$-\infty$	$-\infty$	$-\infty$

Chromosome for the 3rd order model:

Upper bounds	3	3	0	0	0	0	0	0
	nZ	nP	z1	z2	z3	p1	p2	p3
Lower bounds	1	1	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$

Fig. 25 - Genetic Algorithm (GA) coding strategy for 2nd and 3rd order model.

Fig. 25 also presents bound limits for the parameters. As stated before, complex poles and zeros only will appear as conjugated pairs, for an odd order transfer function, real zeros and poles must also be odd. Therefore, in a 3rd order transfer function, there will be one or three real zeros and only one or three real poles. For example, if the GA is searching for solutions for the active power third-order transfer function, whenever $nZ=1$ and $nP=1$, the solution must be decoded as shown in (3.2)):

$$\begin{aligned}
 \text{chromosome} &= [nZ, nP, z_1, z_2, z_3, p_1, p_2, p_3,] \\
 \text{If } M=3, nZ=1, nP=1 \text{ Then} \\
 \theta_p &= [z_1, z_2 + jz_3, z_2 - jz_3, p_1, p_2 + jp_3, p_2 - jp_3]
 \end{aligned} \tag{4.12}$$

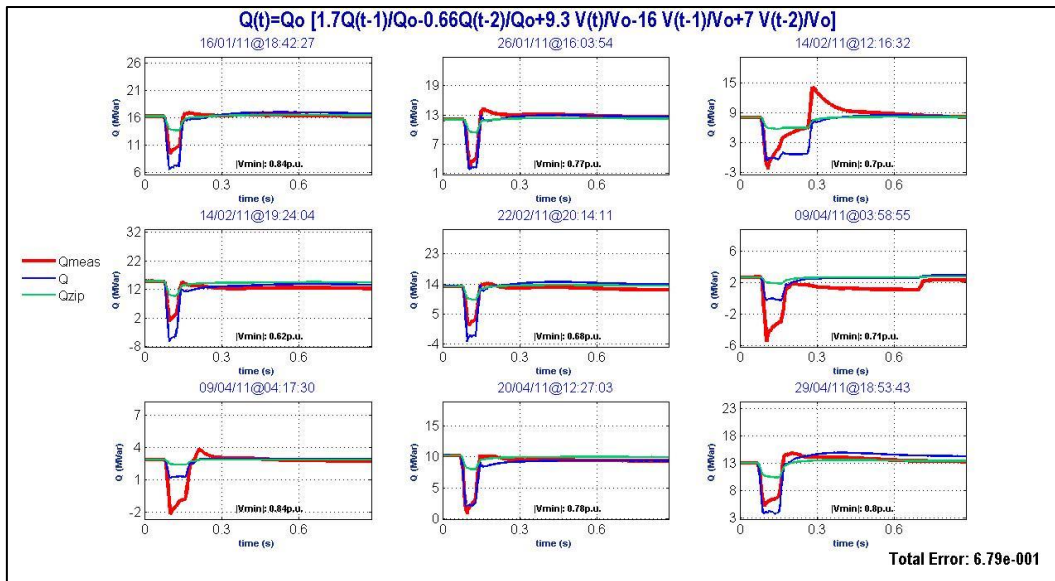
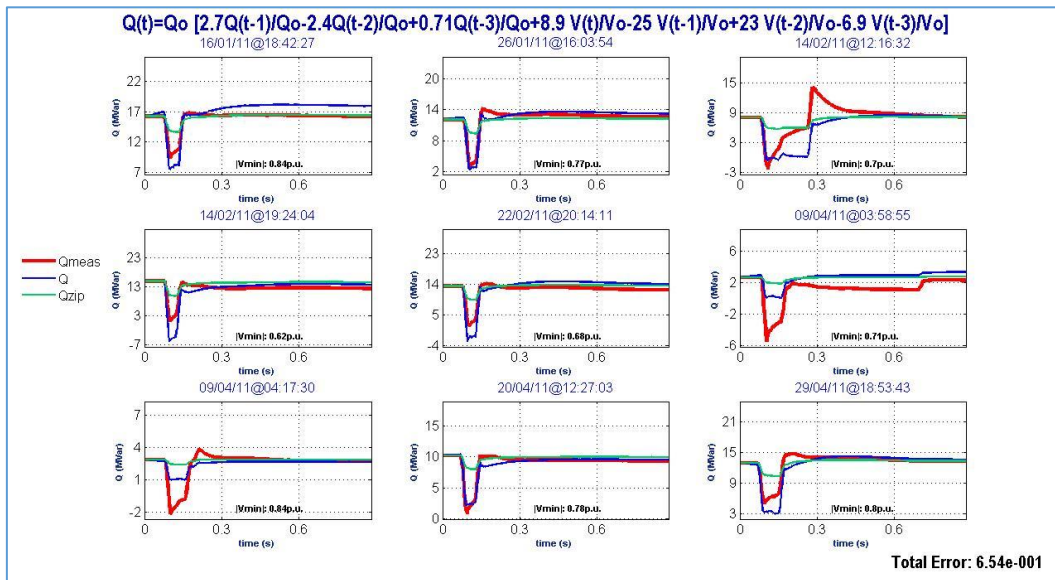
If nZ and nP both are equal to 3, z_3 and p_3 would no longer be the imaginary parts of the complex conjugate zeros and poles, but they would be real zeros and poles respectively. Because complex solutions (zeros and/or poles) always appear as a conjugate, in the second-order model nZ and nP are restricted to be zero or two.

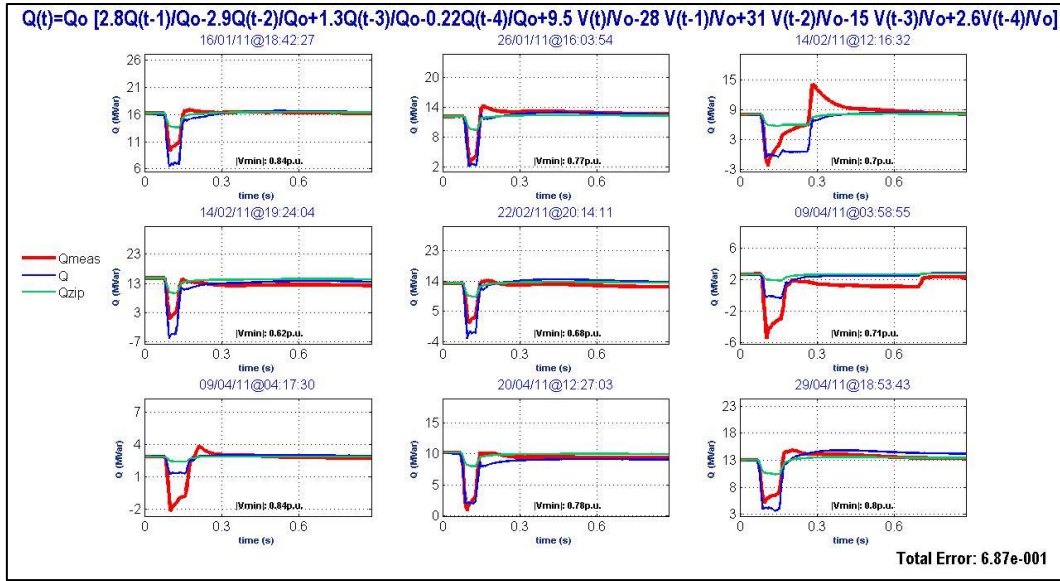
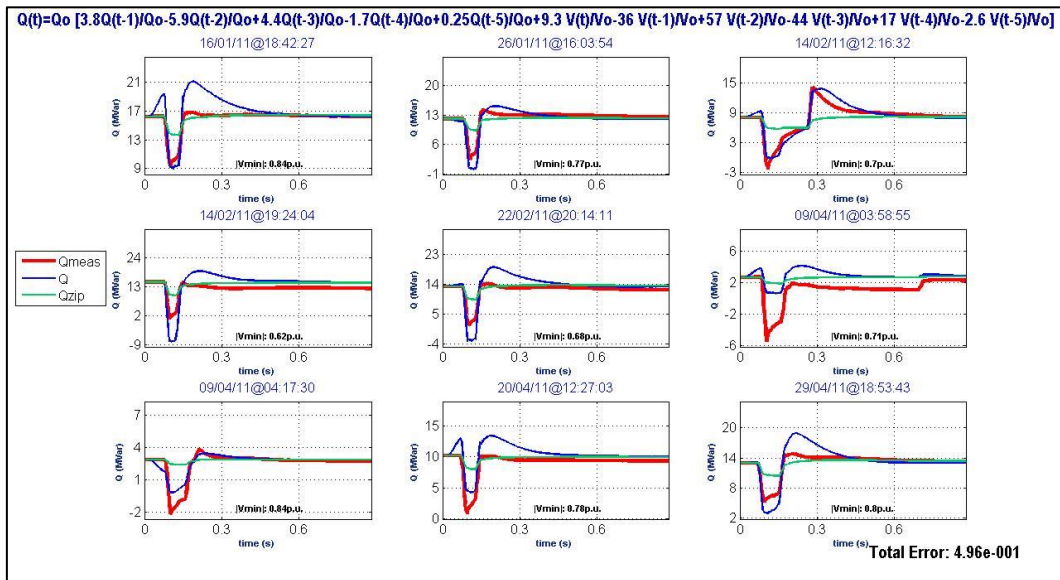
This coding is very flexible, because the heuristic may try a wider scope of solutions, and although there is no guarantee to find the global optimal solution, it is possible to search for real and complex solutions by only generating real negative numbers for $z_{1,2,...}$ and $p_{1,2,...}$. GA-based heuristic was implemented in Matlab using its optimization toolbox and custom crossover and mutation functions were written to respect lower and upper bounds, and these coding characteristics.

4.4.2.

Higher order investigation and a comparison to induction motor different orders

Load induction can be formulated with different degrees of details, which means that the equations may or may not neglect dynamics within this load model, which would result in a more or less complex load model. Common sense may suggest that the higher order a transfer function is, the higher the accuracy to fit model response to measured data. The hypothesis is that the load model complexity (3.1) should be as small as possible. This was one of the main reasons for choosing a second order transfer function, but in this section, the results of an estimation procedure set to the same number of generations, but increasing load model order, are presented below.

Fig. 26 - 2nd order transfer function for reactive power load modelFig. 27 – 3rd order transfer function for reactive power load model

Fig. 28 – 4th order transfer function for reactive power load modelFig. 29 - 5th order transfer function for reactive power load model

Recalling that the induction motor load model is a very common choice for a dynamic simulation implemented in stability software, and also remembering that very often stability software has component library induction load motor built-in within it. Dynamic simulation users may find different induction load motor formulations, regarding its order, and it should be kept in mind that neglecting some types of reactance may lead to formulate the induction load motor as first, third or fifth order [38], [39].

In addition, as it could be seen in the figures above, the comparison between different transfer function orders, after 300 iterations (GA *generations*) the increasing mathematical structure order did not solve a few issues concerning reactive power phenomena, for example. Although the 5th order mathematical structure presented the best result for reactive load modeling (this was the same conclusion in [38], although it was a remarkable longer convergence computational time as the order grows, and also the number of parameters to be set), it can be noticed that some problems of the 2nd order load model remain. For example, if there is a change of signal during voltage sag, the same parameter set cannot fit measured data with and without this signal change. Furthermore, if there is a large variance of Q_0 , it becomes harder for the same parameter set to fit all measured data. The comparison presented from Fig. 26 to Fig. 29 suggests that increasing model order may not improve accuracy significantly.

4.4.3.

Parameter estimation to describe a set of typical contingencies

Expression (4.7) presents the active power load model estimated for typical contingencies recorded in a 69kV load bus bar, defined here as the training data for the experiment conducted in this substation, whose data are plotted in Fig. 30. The green curves are the measured positive sequence active power data from these 9 different contingencies, which originated upstream the measurement point. The red curves are ZIP load model simulations of the contingencies, whose inputs were the measured positive sequence voltages from each one of the 9 selected events, and the dotted blue curves are the simulated load responses, estimated by the optimization procedure.

$$P(\theta_p, t) = P_0 \left[1.89 \frac{P(t-1)}{P_0} - 0.89 \frac{P(t-2)}{P_0} + 2.56 \frac{V(t)}{V_0} - 5.04 \frac{V(t-1)}{V_0} - 2.484 \frac{V(t-2)}{V_0} \right] \quad (4.13)$$

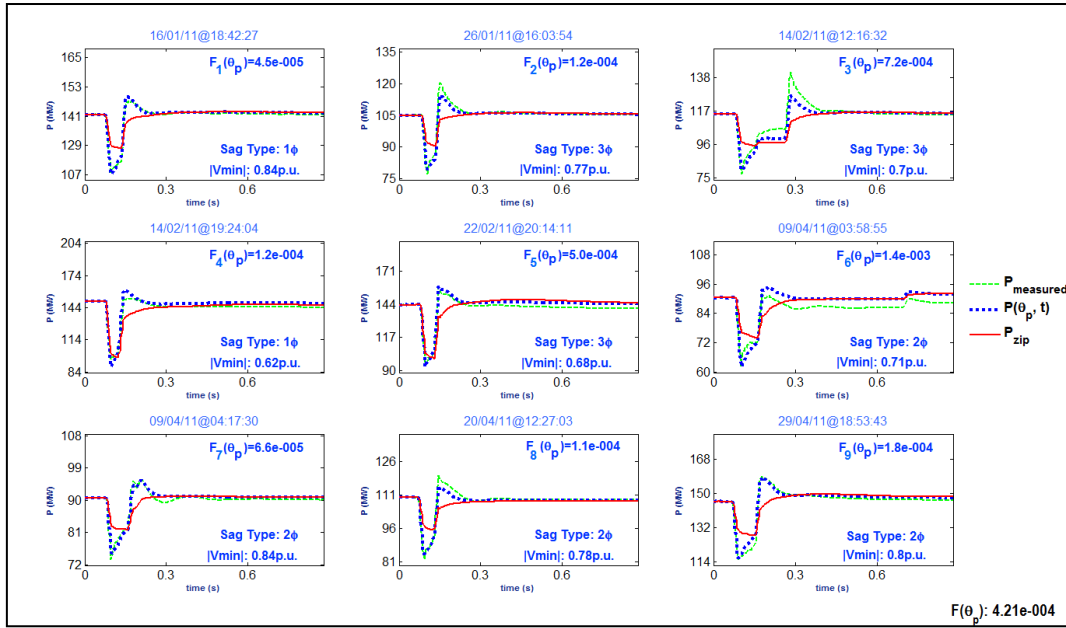


Fig. 30 - Comparison between the measured P (lighter dashed curve), the P simulated by the estimated model (darker dotted curve) and the P described by a generic ZIP model

Although voltage sag data were not shown in the figure, all graphics show the RMS minimum voltage value per unit at the bottom of each graphic, indicating the disturbance severity. Because the transmission utility studied still uses a generic ZIP load model for dynamic simulations, this generic model was plotted in each graphic and subjected to the same voltage sags.

Since the GA has found a parameter set that is able to simulate power behavior during and after the selected contingencies (training data), it should be verified if this solution can represent the contingencies that are separated for validating the proposed model. All of the data chosen for the training data set were recorded in 2011, as shown in the timestamps presented above each graphic. For validating this model, it is essential to use a different dataset, recorded at the same site, to validate the estimated model's capacity to generalize, *i.e.*, to represent the typical behavior of that distribution system for other recorded disturbances.

The contingencies recorded during 2010 were used as the validation data set and they are presented in Fig. 31; note that the same parameter set, formulated in (4.6), that was estimated for the training data, was able to represent different events accurately during and after disturbances originated at the transmission level.

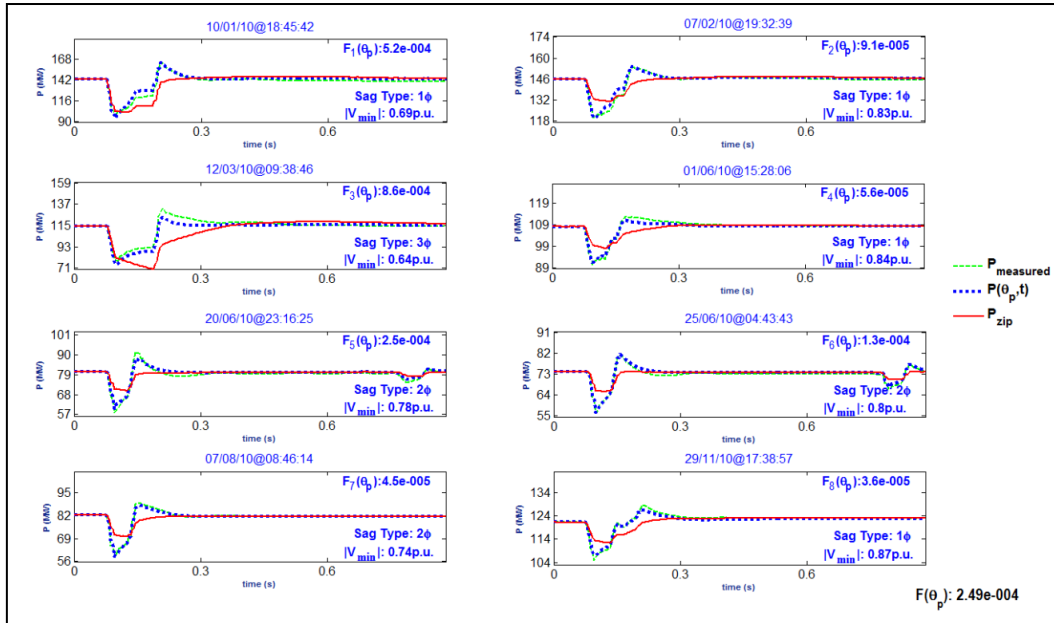


Fig. 31 - Validating the estimated active power model for the 2010 contingencies.

It is more difficult to find a generic load representation for reactive power load modeling because there are devices installed through distribution systems for reactive compensation, such as switched capacitor banks and reactors, and these devices may be or may not be switched on in a given day and may have a completely different configuration in another day. Additionally, this network topology for reactive compensation may be quite different over the course of a single day because of different load demand profiles.

Expression (4.7) presents the formulation to the reactive power load model for typical contingencies, and each curve is disturbed by its respective recorded voltage sag data; although voltage sag data were not shown in the figure, all graphics show the RMS minimum voltage value per unit at the bottom of each graphic, indicating the disturbance severity. Because the transmission utility studied still uses a generic constant impedance load model to represent reactive power in dynamic simulations, this generic model was plotted (solid curves) in each graphic and subjected to the same voltage sags.

Fig. 32 shows that the estimated model could represent records where $Q_0 \sim 30$ Mvar, for both the training data and the validating data (shown in Fig. 33). For those events where Q_0 was higher than 30 Mvar, the estimated model was more pessimistic, and for “low Q_0 ”, the estimated model was more optimistic.

$$Q(\theta_q, t) = Q_0 \left[0.28 \frac{Q(t-1)}{Q_0} - 0.63 \frac{Q(t-2)}{Q_0} + 7 \frac{V(t)}{V_0} + 1.9 \frac{V(t-1)}{V_0} - 8.7 \frac{V(t-2)}{V_0} \right] \quad (4.14)$$

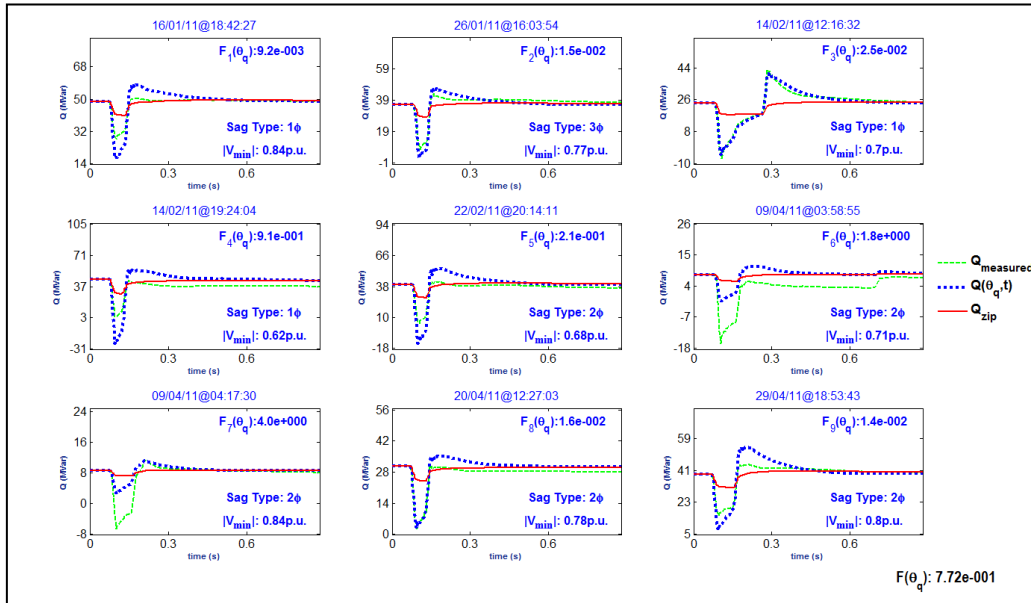


Fig. 32 - Comparison between the measured Q (lighter dashed curve), the Q simulated by the estimated model (darker dotted curve) and the Q described by a constant-impedance load model (solid curve).

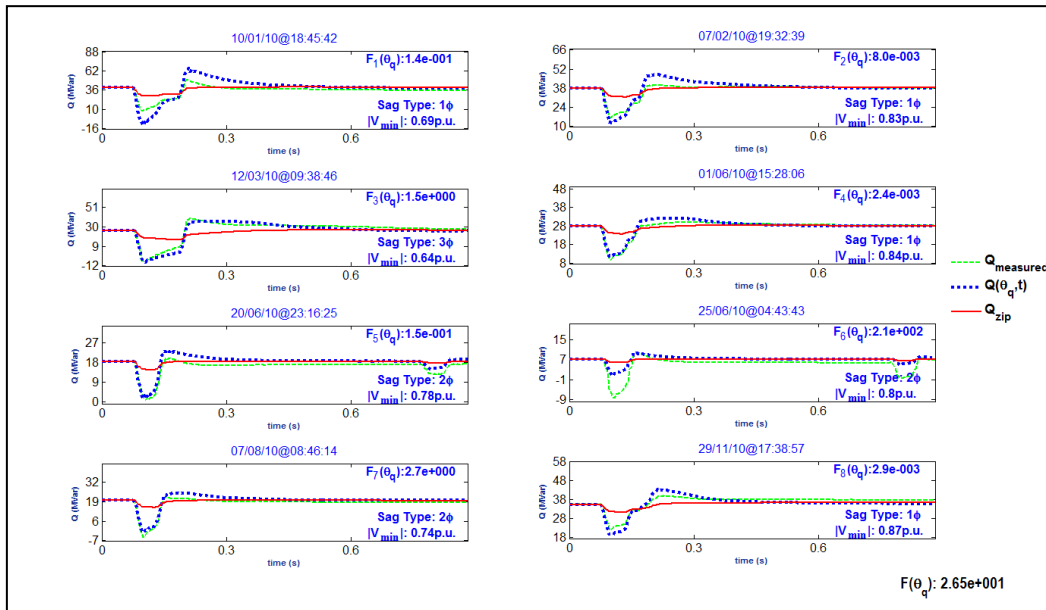


Fig. 33 - Validating estimated reactive-power model for 2010 contingencies.

Another substation is analyzed in Fig. 34 bringing results to the parameter estimation for reactive power load modeling. On the top of the figure, a 2nd order linear transfer function, with the same mathematical structure formulated in equation (4.10), estimated by a GA-based heuristic. This estimated load model calculates all blue curves shown in the figure, subjected to the voltage data for every

contingency, although these measured voltage curves were not plotted in the graphics, but rather summarized by its RMS minimum value during the 1 second window for each contingency recorded.

The red curves plots represent real measurements of reactive power flowing to a distribution system, considering that these 24 contingencies were recorded in different days and different hours of the day.

The green curves are calculated by a Z constant load model, which is the reactive power load model adopted until now in Brazilian dynamic simulations, when there is no information about the distribution system connected to the bulk power system. These curves were drawn as a matter of comparison between the real measurements and the proposed model. All of them were disturbed by the same measured voltage data that disturbed the estimated load model that calculated the blue curves.

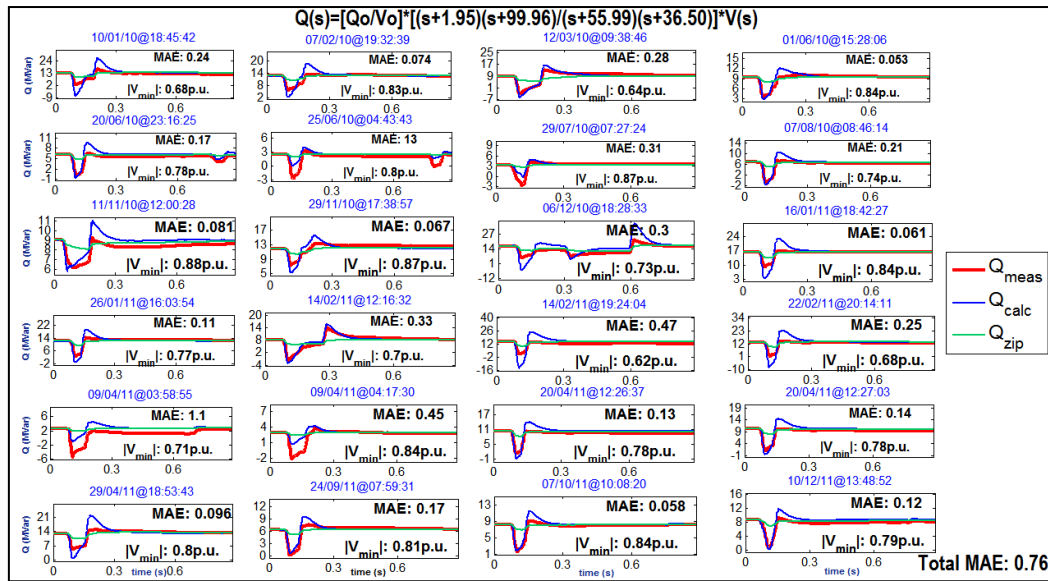


Fig. 34 - A single load model to represent reactive power for 24 different contingencies.

It can be concluded, whether by graphical inspection, or by checking the Mean Absolute Errors (MAE), that a single estimated model was not able to cover as accurately those 24 scenarios. Reactive compensation devices are always spread through the distribution system, and they add complexity to the task of representing reactive flow dynamics.

For the active power load model, Fig. 35 presents graphical results from parameter estimation and the individuals and total MAE showed that a single

estimated model representing active power was able to describe load behavior with good accuracy. It should be reminded that this distribution system has no significant dispersed generation.

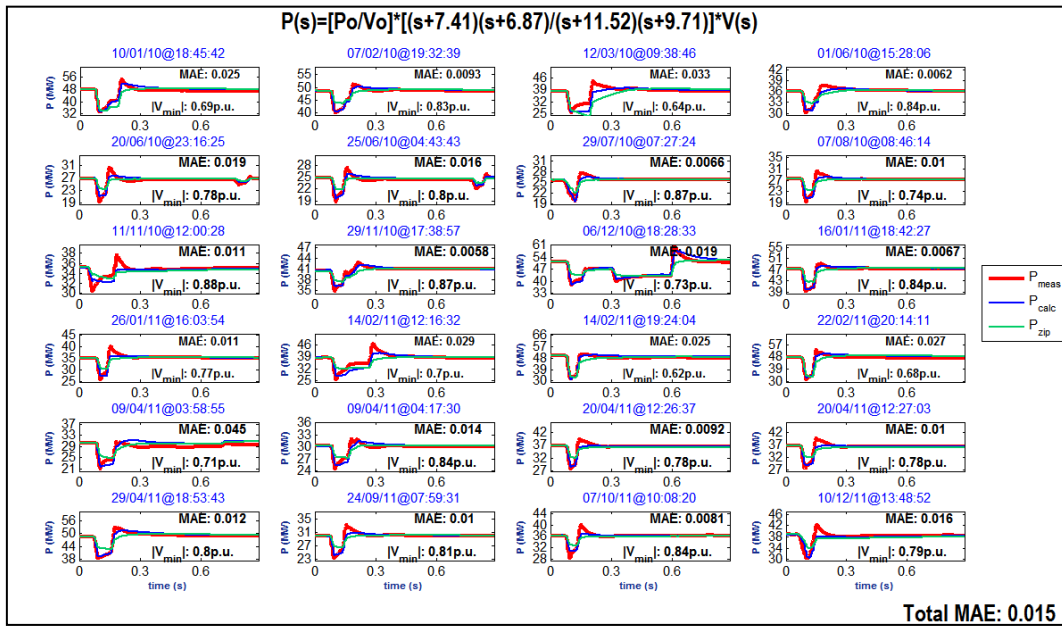


Fig. 35 - A single load model to represent active power for 24 different contingencies.

4.4.4.

Parameter estimation to reproduce one event of interest

The example in this section uses data from a disturbance recorded by an IED at a 69 kV bus bar. According to the utility's operational report, this disturbance was caused by an atmospheric discharge on a 230 kV transmission line near the 230/69 kV transformer and the distribution system contains two thermo-generators with installed capacities of 110 and 30 MW each, as illustrated in the schematic diagram of Fig. 13.

After fault clearance, there is a power oscillation between these generators and the synchronous machines from the bulk power system, as it can be seen in Fig. 36 (red curve). This behavior is quite different from that in distributions systems with negligible or no Dispersed Generation (DG).

To estimate this parameter set, the second-order continuous transfer function, formulated in (4.10), was used, and the constraint imposed on the GA parameter estimation was that only complex conjugated pairs of the poles were valid for the optimization procedure. This constraint was motivated by visual inspection of the

oscillatory dynamics recorded after a disturbance clearance by the protection system (in Fig. 36, around the instant 0.2 sec); recalling that an inverse Laplace transformation of a continuous transfer function, with complex conjugated pairs of the poles, evaluates an oscillatory step response, as illustrated in Fig. 21.

To achieve this solution, it was imposed in the codification shown in Fig. 25 that $nP=0$, resulting in complex conjugated poles with negative parts. The chromosome length was 6. Fig. 36 presents the result of the estimated load model for active power at the analyzed bus. Once again, the lighter dashed curve is the active power measured by the IED, the darker dotted curve is the active power simulated by the proposed model shown at the top of the figure. Finally, the solid curve is a constant impedance model (P_{zip}), which is still used by the utility to represent the load in dynamic simulations when there is no information of the distribution grid topology.

Clearly, the constant impedance model is not able to approximate the load behavior, neither during nor after the disturbance. This generic constant impedance model is an overoptimistic load model, as defined in [4], which means the power deviation is underestimated and may lead to a misevaluation of the stability margins and even to voltage collapse in critical cases.

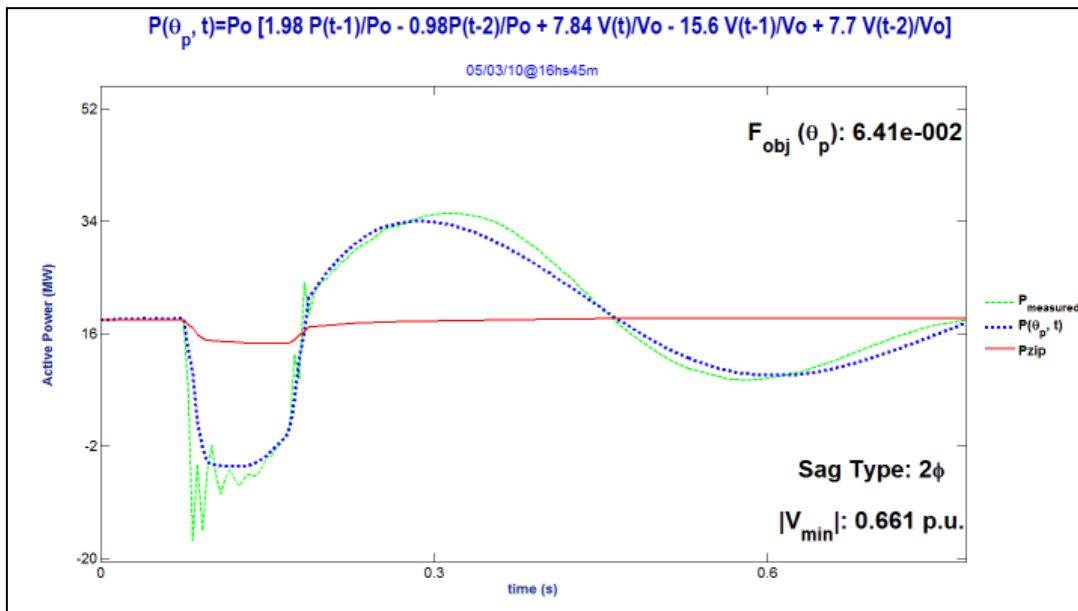


Fig. 36 - Reproducing a specific event: a distribution system containing DG and active power behavior.

Fig. 37 shows the reactive power curve, both measured and simulated, and the parameter set estimated for the reactive power autoregressive model. The

reactive power expression is shown at the top of the figure. Once more, the constant impedance model (Q_{zip}) was not able to represent the load dynamics and the sign change of the reactive power during the fault.

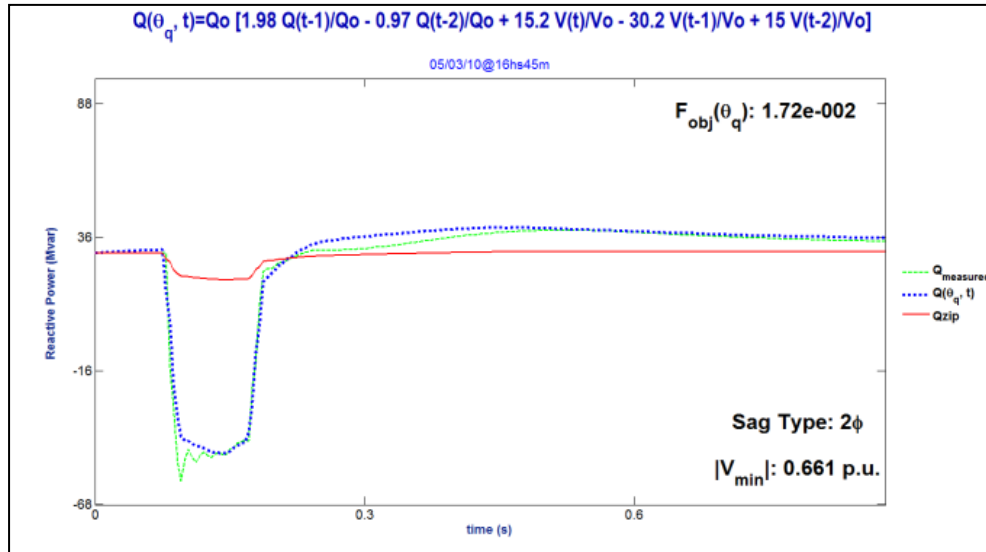


Fig. 37 - Reproducing a specific event: a distribution system containing DG and reactive-power behavior.

4.5. Model Validation

ANATEM is a simulation software of electromechanical transients used for the analysis of dynamic stability, supporting the operation and planning of the Brazilian interconnected power systems. The estimated load model for the distribution system containing DG (shown in the previous section, in Fig. 36 and Fig. 37) and for ten more 69kV substations from the Brazilian transmission/distributions frontier were validated in ANATEM [40]. The entire interconnected national power system is modeled in this software, and the Brazilian transmission system operator (ONS) publishes in their website the input files for this software detailing the various components of the Brazilian electrical system at 750kV, 500kV, 230kV levels. The 69kV distribution systems, however, are not represented in further details, but they are aggregated in equivalent ZIP models.

To reproduce the event shown in Fig. 36 and Fig. 37, a short-circuit was applied at the 500 kV level (transmission system) and after 80 ms, the short circuit was eliminated and a 500 kV transmission line was opened. The ZIP models used

for comparison are those published by the ONS, corresponding to P_{ZIP} : 60% of constant P and of constant Z, while Q_{ZIP} is 100% constant Z.

Fig. 38 and Fig. 39 respectively present the active and reactive power curves simulated in ANTEM, at the same 69 kV bus containing DG of Fig. 13; dashed curves are the active and reactive power evaluated with measured-based load models implemented in the simulation software and solid curves are the active and reactive power evaluated with generic ZIP load models implemented in the simulation software.

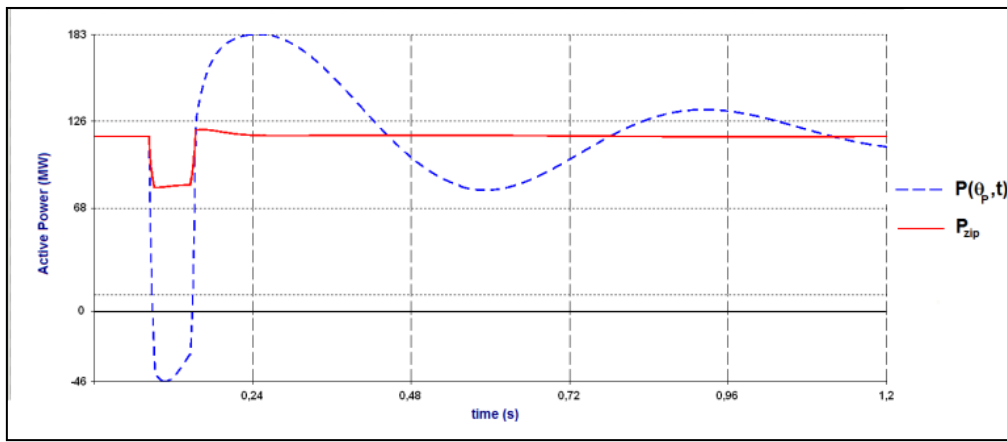


Fig. 38 - Active power simulated at the 69 kV bus bar with DG by a constant-impedance load model (P_{zip}) and the last section's estimated model (P).

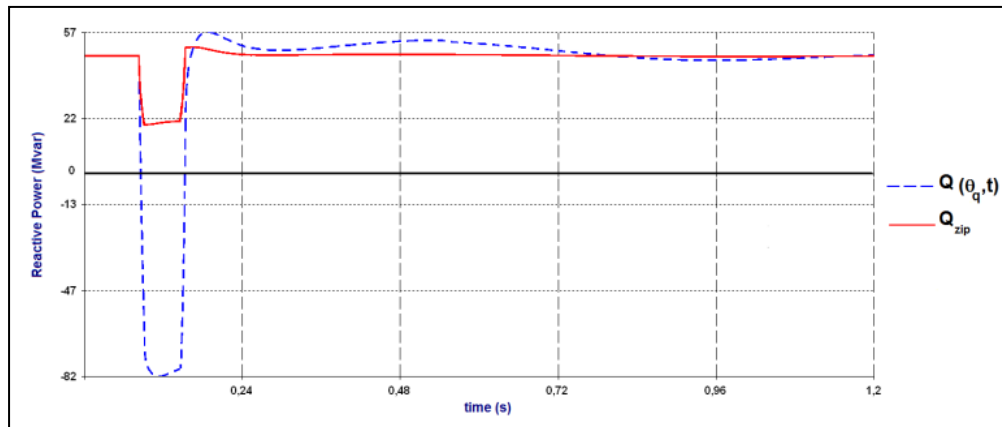


Fig. 39 - Reactive power simulated at the 69 kV bus bar with the constant-impedance (Q_{zip}) and estimated models (Q).

Power oscillation modes could not be represented by the constant impedance model. In addition, both the active and reactive power flows have changed signs during the disturbance, exactly like the real measurements have shown in Fig. 36

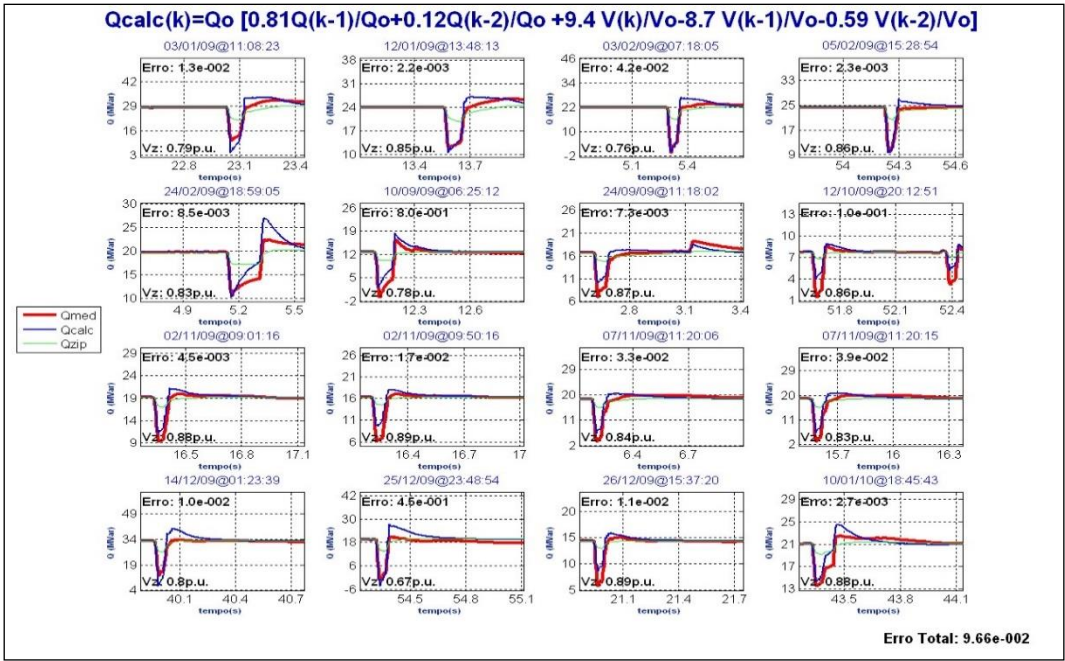


Fig. 42 – Q_{med} is the reactive power measured, Q_{calc} is the estimated load model, described by the equation at the top of the figure, and Q_{ZIP} is 100% constant Z (substation 1)

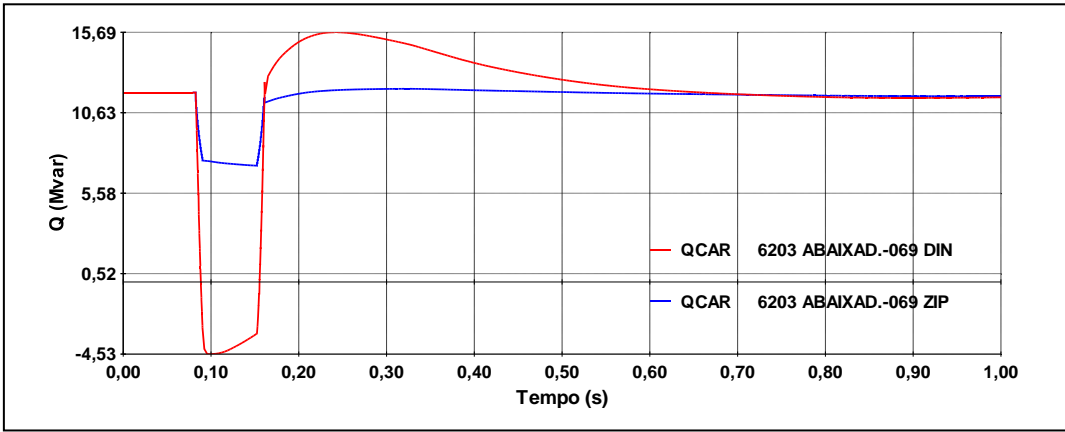


Fig. 43 - Comparison between Q_{ZIP} and the estimated load model of Fig. 42 (substation 1) implemented in ANATEM.

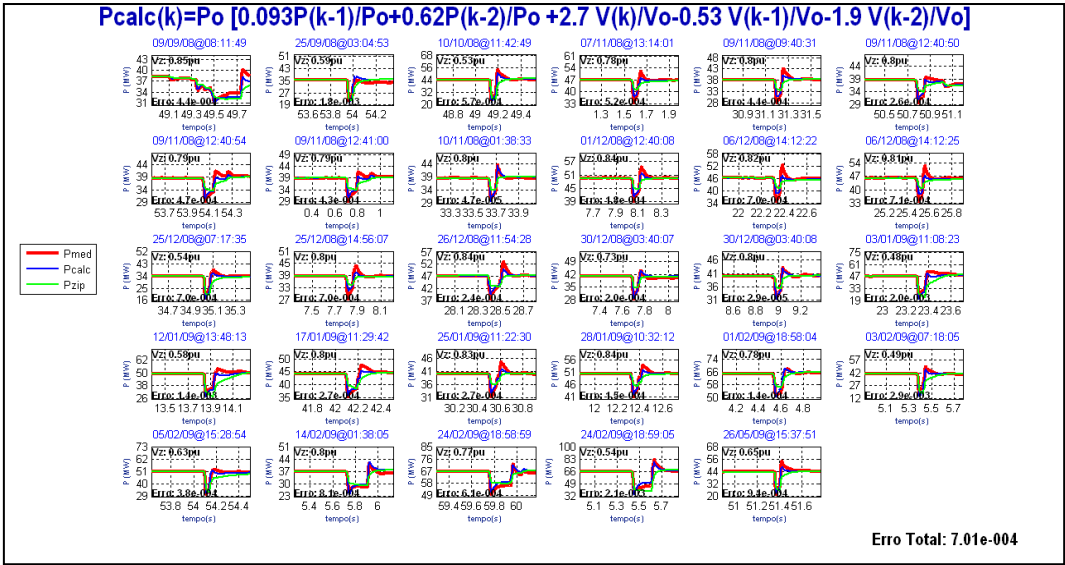


Fig. 44 - P_{med} is the active power measured, P_{calc} is the estimated load model, described by the equation at the top of the figure, and P_{ZIP} is 60% constant P and 40% constant Z (substation 2).

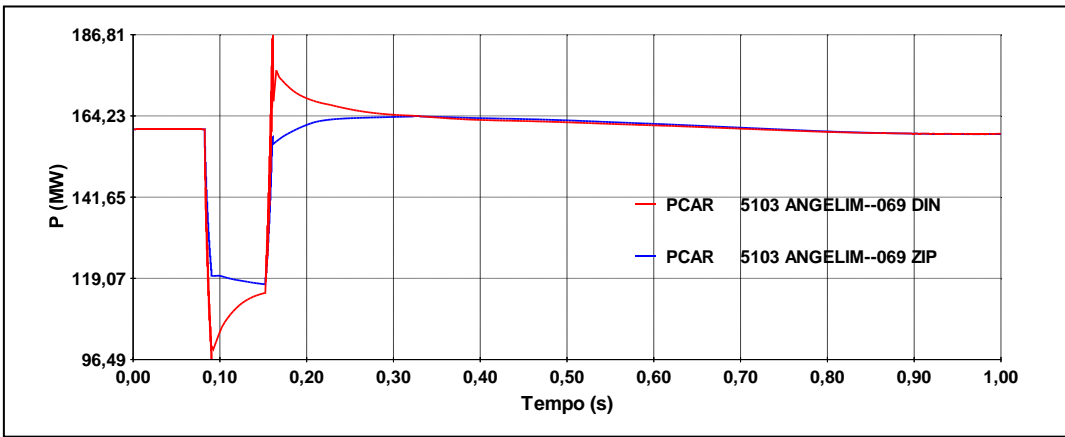


Fig. 45 - Comparison between P_{ZIP} and the estimated load model of Fig. 44 (substation 2) implemented in ANATEM.

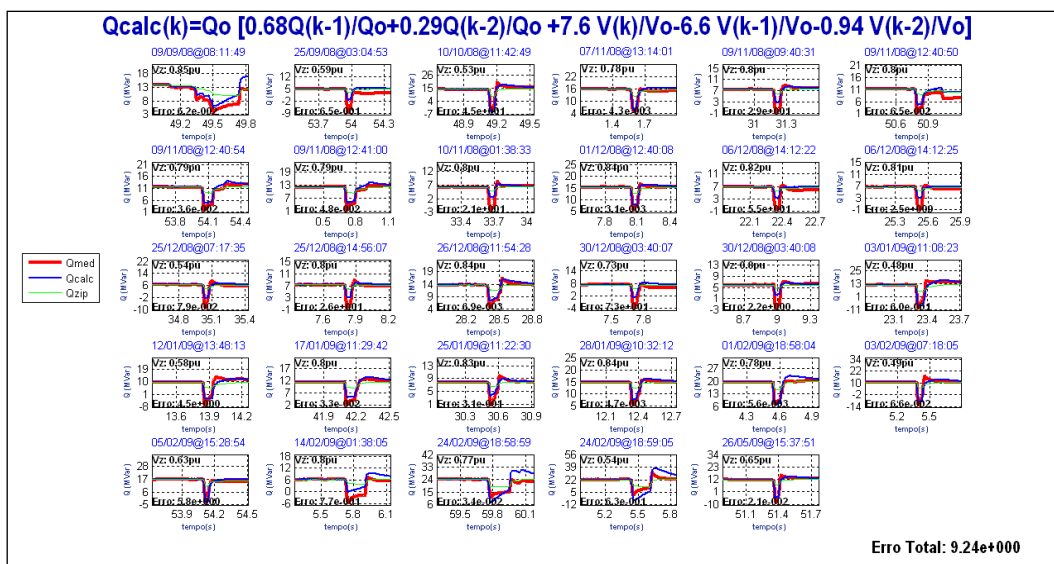


Fig. 46 - Q_{med} is the reactive power measured, Q_{calc} is the estimated load model, described by the equation at the top of the figure, and Q_{ZIP} is 100% constant Z (substation 2).

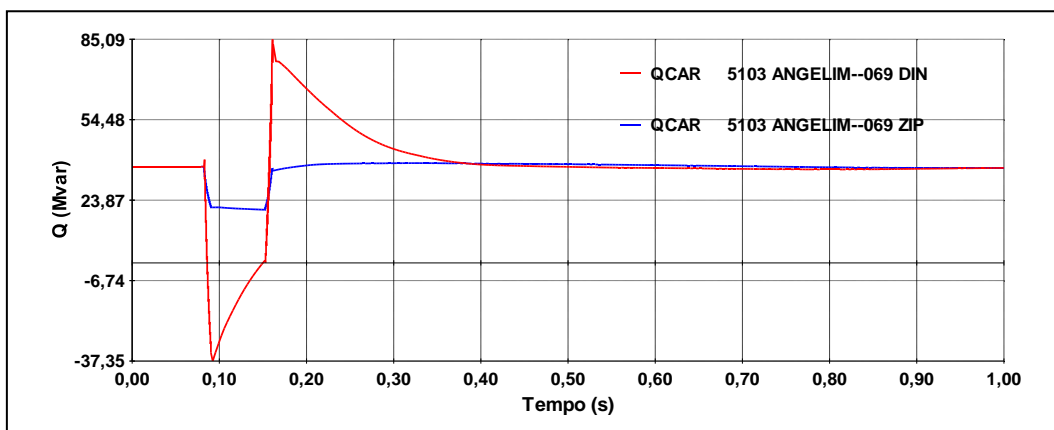


Fig. 47 - Comparison between Q_{ZIP} (constant Z) and the estimated load model of Fig. 46 (substation 2) implemented in ANATEM

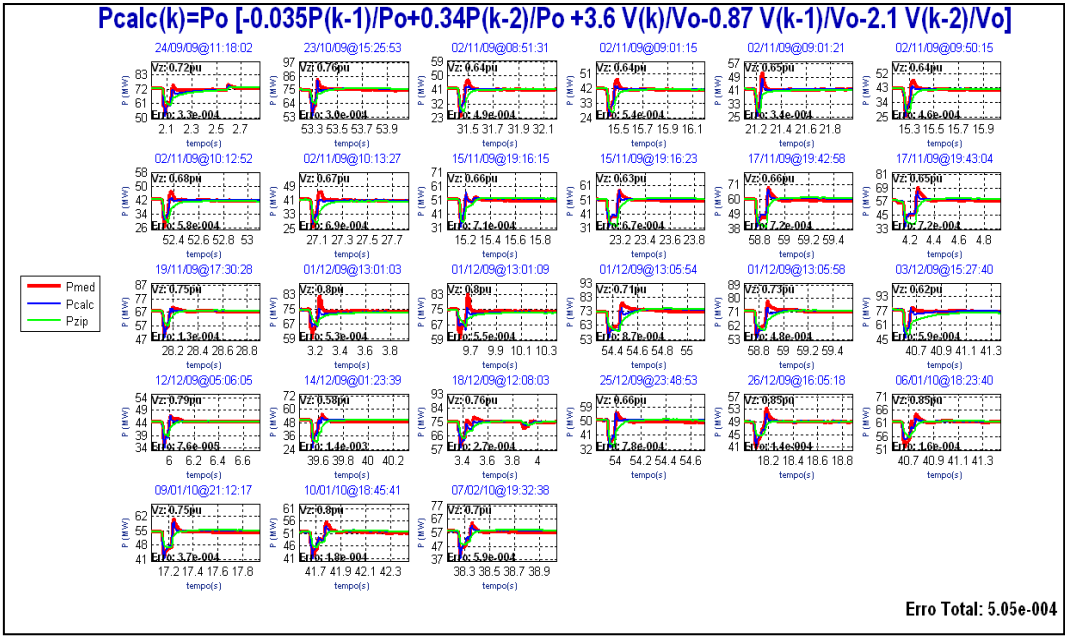


Fig. 48 - P_{med} is the active power measured, P_{calc} is the estimated load model, described by the equation at the top of the figure, and P_{ZIP} is 60% constant P and 40% constant Z (substation 3)

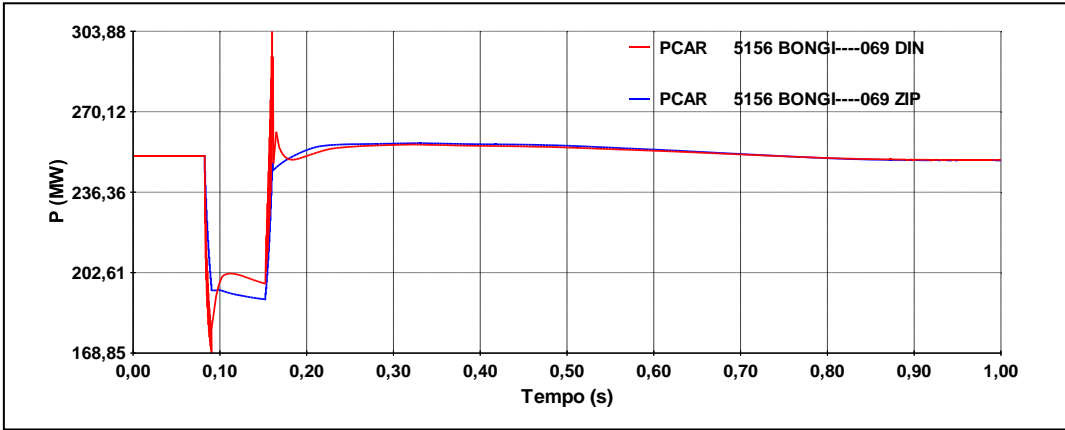


Fig. 49 - Comparison between P_{ZIP} (60% constant P and 40% constant Z) and the estimated load model of Fig. 44 (substation 3) implemented in ANATEM.

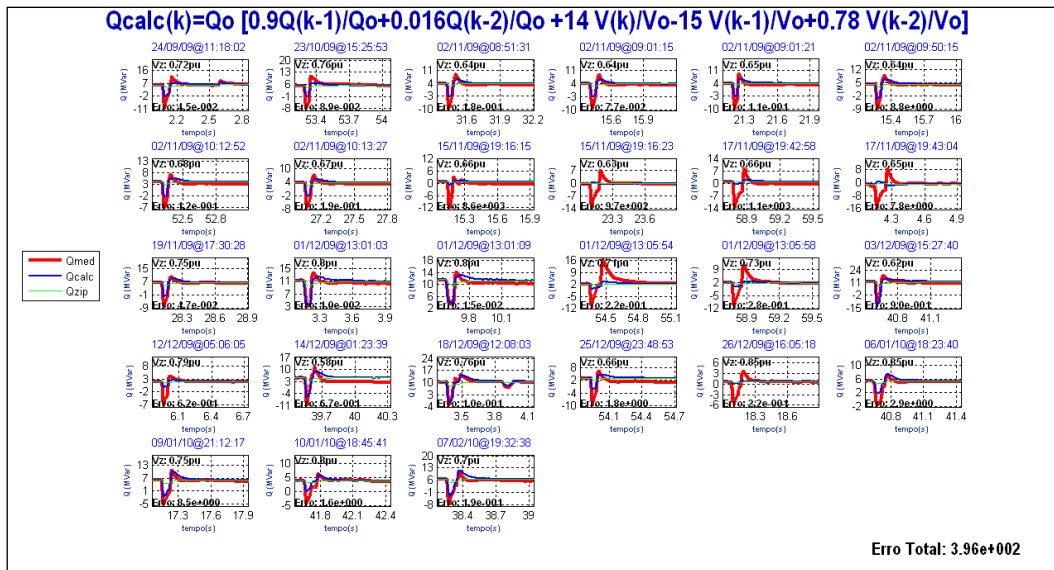


Fig. 50 - Q_{med} is the reactive power measured, Q_{calc} is the estimated load model, described by the equation at the top of the figure, and Q_{ZIP} is 100% constant Z (substation 3).

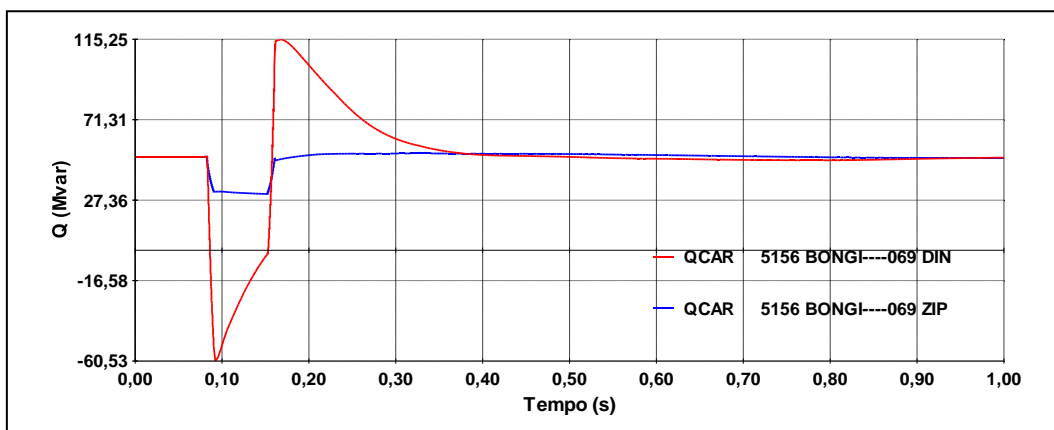


Fig. 51 - Comparison between QZIP (constant Z) and the estimated load model of Fig. 40 (substation 2) implemented in ANATEM

4.6. Summary of the chapter

The load response to large system disturbances has been investigated using field data measurements. The following conclusions are listed below:

- The use of traditional static load models should be replaced for dynamic load models in simulations of system stability. It was possible to see that static load models underestimate the load response to disturbances whenever the voltage dropped below 0.9 p.u.

- The load-to-voltage dependency has been presented in examples recorded in every month of a year, thus it represents a critical aspect for the design of security margins, and optimistic models such those found in the ANATEM files of the Brazilian interconnected system, may lead to underestimate the assessment to system security. Particularly for the reactive power results, it could be noticed that constant impedance load model is frequently over-optimistic regarding the real deviations in the reactive power, including some transitory flux inversions that the constant impedance relation is not able to represent mathematically.

- It has been presented an application of the load modeling methodology described in chapter 3. The adequacy of the dynamic load model has been investigated for different recording among CHESF frontier to regional distribution systems. It has been shown that the second order transfer function representation was able to estimate parameter sets that could describe load behavior accurately.

- The validation of this measurement-based load models was clearly better in active power representation than in reactive power representation. One of the reasons is that reactive compensation is spread in the studied systems, but they vary according to system operation conditions and there was no information whether capacitor banks are on or off.

5

Load Modeling for Conservation Voltage Reduction – A Case Study

Conservation Voltage Reduction (CVR) is not a new technique [41]-[44], but recently this concept has received a renewed interest aiming to quantify its benefits. CVR also seems to be tuned with the new global order of developing smarter solutions, aiming to improve energy usage efficiency. Network operators have always tried to optimize the network energy transfer, developing more flexible ways of managing electric grid resources. CVR has been investigated through field-data analysis researches [45]-[49], and this Volt-Var optimization application has been considered as an alternative low-cost approach to maximize the electricity network capacity is energy conservation through voltage reduction during load consumption peak times, therefore improving the efficiency of the network assets utilization.

The main idea of CVR relies on the positive correlation between voltage and active power, assuming that for a small reduction within acceptable limits in voltage at the customer service entrance, there is a proportional small reduction in load demand, yielding an energy consumption reduction while the voltage remains reduced. Therefore, accurate load models are needed to support decision making in whether it is worth to implement this technique or not, in other words, can demand response, caused by dynamic voltage regulation, reduce the peak demand of a primary substation? How much energy can be saved applying this technique?

All around the world, there are standardization on voltage range acceptable limits, and EN 50160 establishes in Europe that voltage magnitude must be delivered within $\pm 10\%$ and, depending on voltage level, the range must be within $\pm 5\%$ [50]. It has always been an usual recommended practice to maintain the voltage at the higher-end of the acceptable range, but the CVR approach proposes to maintain the voltage at the lower-end of the acceptable range, in compliance to voltage standardization, during selected times of the day, in order to save energy, lower peak demand and losses. It is known that the relationship between voltage

and power depends on the load composition, and the goal of this case study is to quantify this and try to associate it to the load mix classes previously identified.

Some indicators were established to quantify CVR benefits and the most general way to define the CVR factors is:

$$\begin{aligned} CVR_{f-P} &= \frac{\Delta P\%}{\Delta V\%} \approx Kp \\ CVR_{f-E} &= \frac{\Delta E\%}{\Delta V\%} \\ CVR_{f-Q} &= \frac{\Delta Q\%}{\Delta V\%} \approx Kq \end{aligned} \quad (5.1)$$

Which means that CVR_f aims to quantify the ratio between power reduction (or energy saving) and voltage reduction. It is also possible to quantify reactive power reduction in this way, although it is far more difficult to scrutinize this relationship.

The data analysis approach will combine a set of techniques to estimate the CVR factors from a huge database of measurements. It will include:

- **Automatic voltage reduction detection**
- **Exploratory Data Analysis (EDA).**
- **Parameters set estimation**
- **Statistical inference.**

In section 5.1, it will be shown how EDA was used as a preliminary data analysis supporting the stratified sampling scheme. It was a very intuitive tool to separate data into similar scenarios, regarding different time frames (seasons, hours of the day, weekdays/weekends). EDA is a data analysis approach that relies on graphical interpretation of data patterns, allowing the data to reveal its underlying structure [51]. The seminal work in EDA is [52] and over the years it is considered the most intuitive way to previously get insights about the data under investigation. In the same section, the automatic voltage reduction detection algorithm, in need to determine the exact timestamps of the voltage reductions used to evaluate the CVR factors, is explained and also the smoothing technique to extract noise, outliers and missing data from voltage and demand signals. Section 5.3 recalls the exponential recovery load model, presenting a short analysis of the model step response that visually justifies its choice as a CVR assessment tool. Parameter estimation and the

stratified sampling scheme are detailed in section 5.4, and finally in section 5.5 the results of CVR factors for the different substations under study are supported by statistical inference of their confidence intervals.

5.1. Case Study Description

This case study was an application of load modeling to estimate the effect of CVR in some British distribution systems, whose data measurements were collected from trials conducted by Electricity North West (ENWL) within the Customer Load Active System Services (CLASS) project [53] from June 2014 up to May, 2015, in England. The CLASS project main goal was “to demonstrate an innovative, low cost and easily deployable solution to provide active voltage management for demand response capabilities and network voltage regulation services”.

In the context of the CLASS project solution, the core functionalities were configured in the Autonomous Substations Controller (ASC). The voltage reductions were executed by the ASC, programmed to control voltage at Primary Substations, and also to deliver automatic frequency response, according to power systems operational needs. The ASC interfaces through a SCADA infrastructure to an Automatic Voltage Control (AVC) scheme. The voltage steps were performed by the AVC acting on the taps of the On Load Tap Changing transformers (OLTC) of the primary substations, both connected in parallel. Frequency, power factor, voltage, active and reactive power data were sampled once every second by Remote Terminal Units (RTU), and sent to a web server.

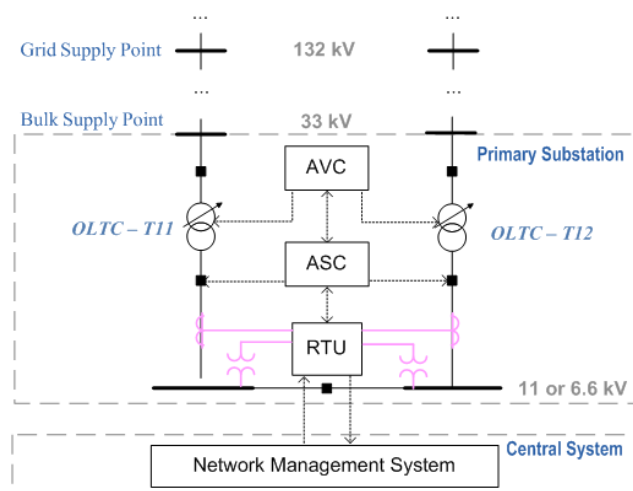


Fig. 52 - Schematic Diagram of primary substation monitoring

For the purpose of evaluation of CVR indicators from the substations in the ENW area, it was designed a set of scheduled trials, covering all days of the week, all the hours within each day, in different seasons. The methodology of the substations selection is described in details in [55], and a test scheme was designed for each primary substation, with representative periods for voltage reduction/elevation, during 1-year period.

The selection of primary substations for the CLASS project went through a process of analysis and the final selected ones aim to include a set of load types that can be considered representative of ENWL network, where hopefully the conclusions estimated from these substations data could be applied to other similar substations. Among the main goals of these scheduled trials, the results that will be presented next were focused on the investigation of the estimation of voltage-demand relationship, regarding different load types, in order to quantify demand response to voltage changes during peak and off-peak hours. Fig. 53 presents a table from [56] detailing the tests conducted for each of the project research targets and T1 and T2 were studied in this thesis.

An outline of the CLASS Trials is presented below.

Table 1 - List of CLASS Trials

Reference	Description	Objective	Technique	Trial Period	Customer Survey Requirement
T1	Load modelling	Establish voltage-demand relationship	Raise and lower tap positions	Across entire annual cycle	No
T2	Peak demand reduction	Demand response for peak reduction	Lower tap position	Peak demand	Yes
T3a	Stage 1 frequency response	Response to reduce demand	Switch out transformer	Anytime	Yes
T3b	Stage 2 frequency response	when system frequency falls	Lower tap position	Anytime	Yes
T4	Reactive power absorption	Reduce high volts on transmission network	Stagger tap position	Minimum demand	No

Fig. 53 - CLASS trials, extracted from [56]

For the purpose of evaluation of CVR indicators from the substations in the ENWL area, there was designed a set of scheduled trials, covering all days of the week, all the hours within each day, in different seasons. The methodology of the

substations selection is described in details in [55], and a test scheme was designed for each primary substation, with representative periods for voltage reduction/elevation, during 1-year period, starting in June 1st, 2014 up to May, 31st, 2015.

The substations selected for the trial period are divided into three previously identified categories:

Type 1 - a load mix composed mainly by industrial/commercial consumers,

Type 2 - a load mix composed mainly by residential consumers,

Type 3 - a mix of both categories, where there is a roughly identical share of residential and non-residential load mix.

To give the reader a first glimpse of the substations` characteristics, the approach used is inspired in Exploratory Data Analysis (EDA), which is a philosophical approach that is intended to let the data reveal patterns through visual inspection. Fig. 54 and Fig. 55 show the daily profiles of a Type 1 substation, Avenham, respectively on January/2015 (winter) and July/2014 (summer). The y-axis registers the active power demand, whereas the x-axis refers to the hours of the day. It can be observed the typical bell-shaped curve of load profile every weekdays and weekends. The main difference between winter and summer are the magnitudes of peak demand.

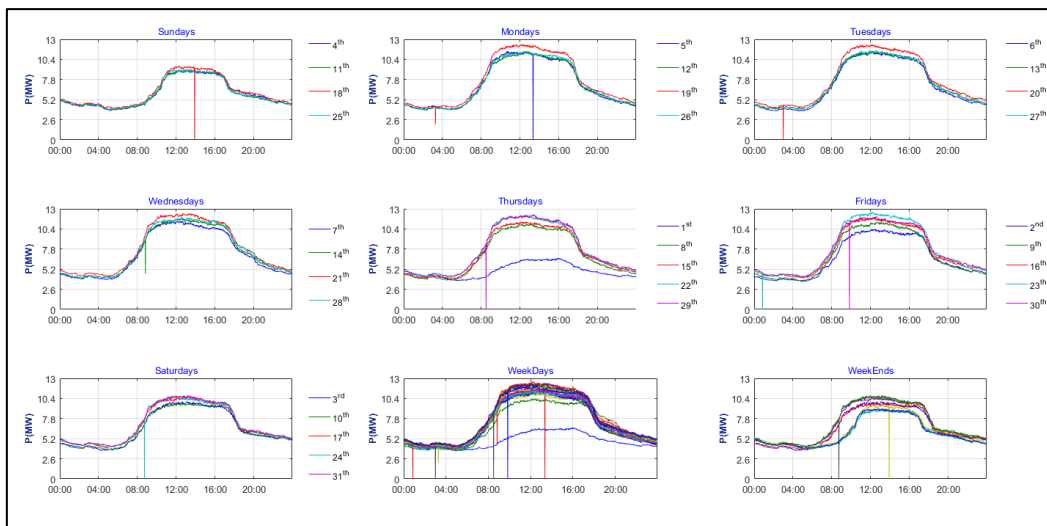


Fig. 54 - Daily Profile of Avenham (Type 1), January/15 (Winter)

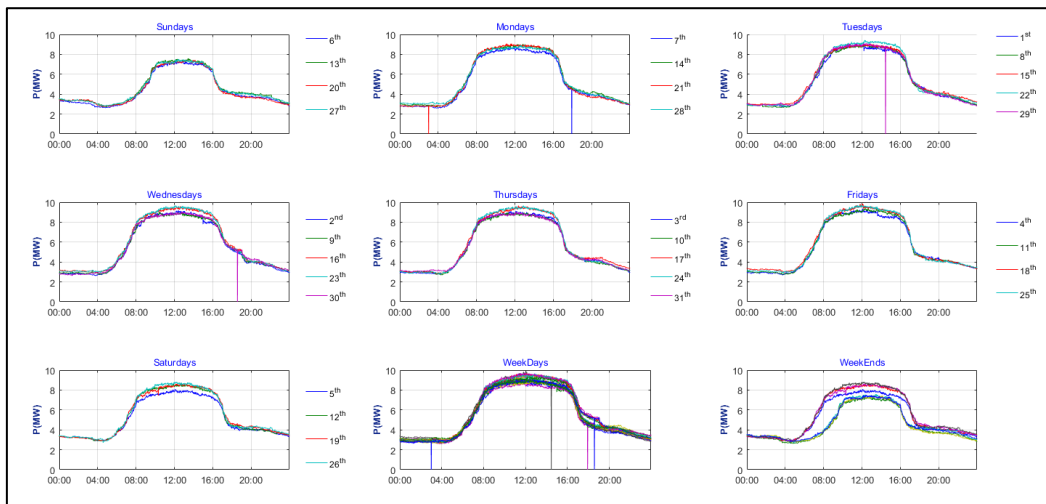


Fig. 55 - Daily Profile of Avenham (Type 1), July/14 (Summer)

Fig. 56 and Fig. 57 show the daily profiles of a Type 2 substation, Fallowfield, respectively on January/2015 (winter) and July/2014 (summer). It can be observed the shape of a residential load profile, characterized by the distinguished light, medium and heavy load during the days. The peaks in summer are much smaller than in winter, and the weekend profiles are very similar to weekdays, in both seasons.

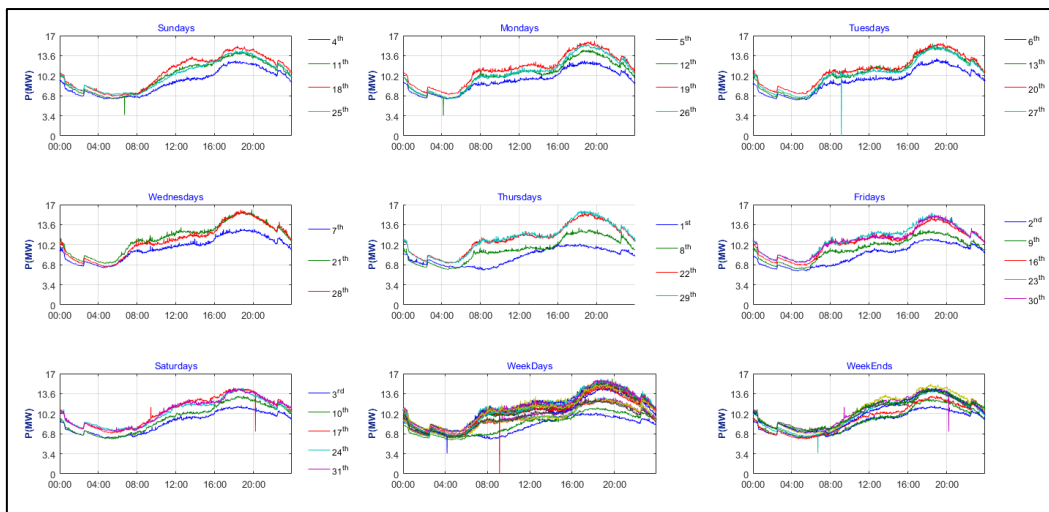


Fig. 56 - Daily Profile of Fallowfield (Type 2), January/15 (Winter)

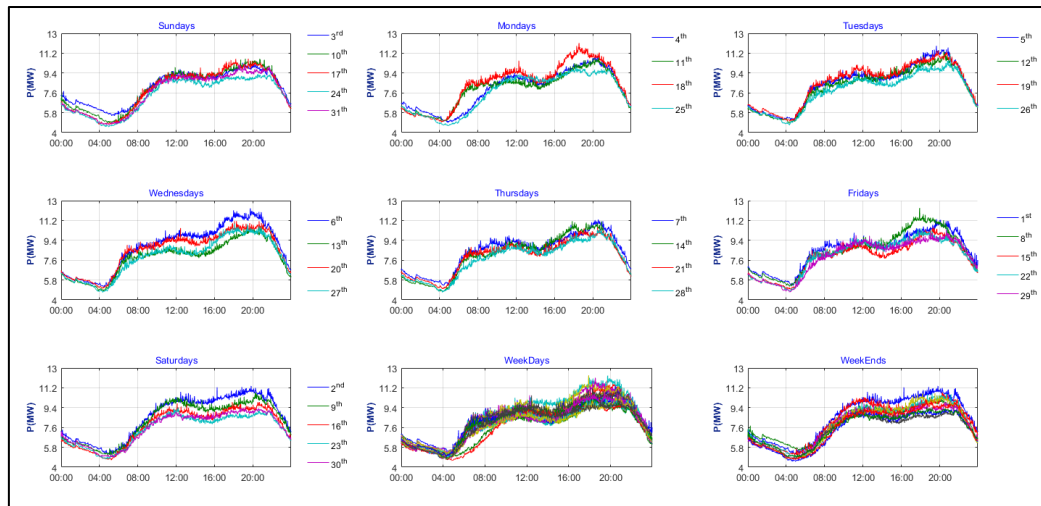


Fig. 57 - Daily Profile of Fallowfield (Type 2), May/15 (spring)

Fig. 58 and Fig. 59 show the daily profiles of a Type 3 substation, Kitt Green, respectively on January/2015 (winter) and May/2015 (spring). The shape of the load profiles in winter looks like a residential load mix, whereas in spring it looks more like the bell-shaped commercial/industrial load mix. The peaks in spring are much smaller than in winter, and the weekend profiles are very distinct from weekdays, in both seasons.

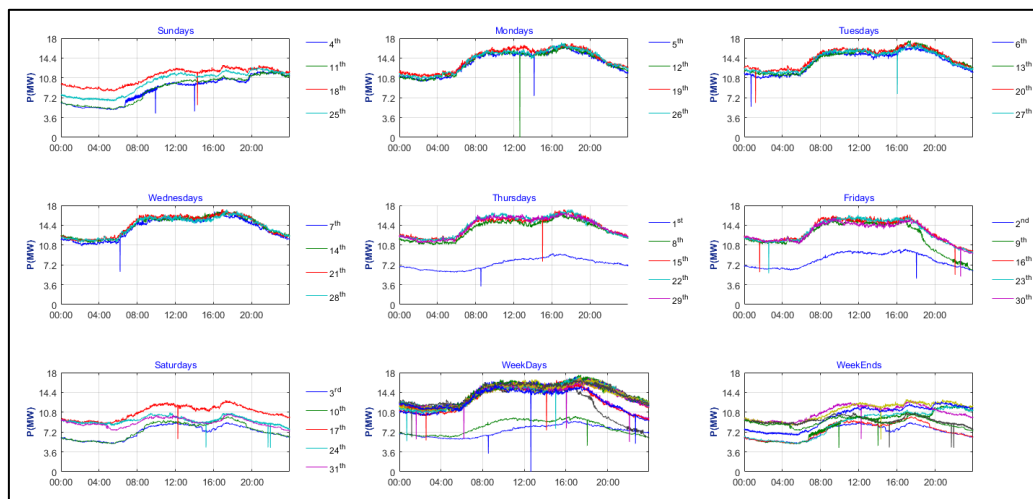


Fig. 58 - Daily Profile of Kitt Green (Type 3), January/15 (Winter)

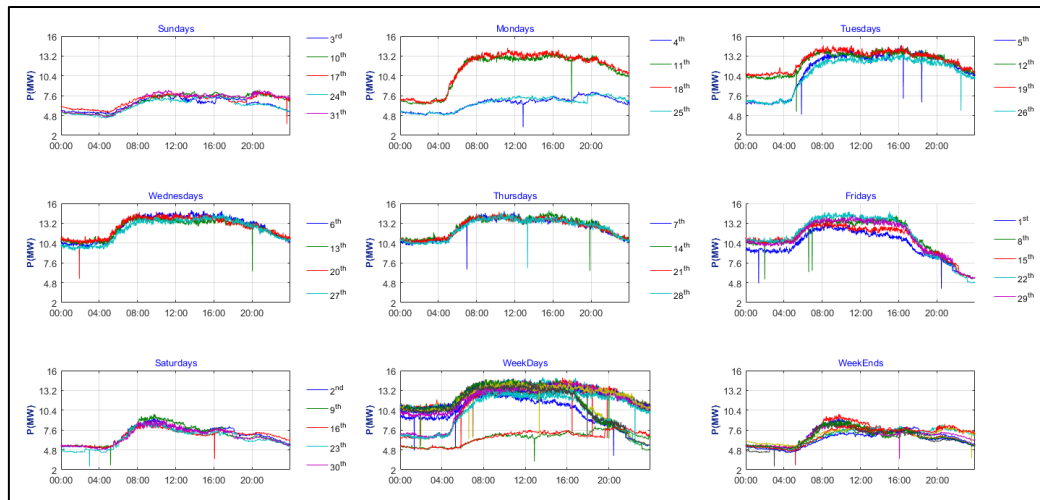


Fig. 59 - Daily Profile of Kitt Green(Type 3), May/15 (Spring)

These daily load profiles are very useful to identify *outliers*, which is an important statistics concept, that represents a single data value or a data set that is remarkably different from data observed within an investigation. In this context, a daily profile that could be considered an outlier is the profile observed in a holiday or weekend that are quite different from an ordinary weekday. Fig. 54 and Fig. 56 show daily profiles from January/2015, and January 1st stands out, and it is obviously one of the 365 days that should be analyzed separately, but the last obvious May 4th and May 26th, both Mondays are the daily profiles presented in Fig. 57 and Fig. 59, although in Fig. 57 the difference from the other Mondays is much more subtle than in Fig. 59.

Missing data and abnormal operational periods can be identified in daily profiles. From Fig. 54 to Fig. 59 there are several missing data records, making the line series look as if they are “melting” in some moments of the day. Abnormal operation periods can be noticed in Fig. 60, in which the daily profiles were recorded from one of the two parallel transformers in Kitt Green substation (from June/14 to October/14 the data available from all substations were recorded from one of the two parallel transformers). So, in Fig. 60 the transformer is out of service during a few minutes after 8am up to a few minutes after noon, on Monday 21st, while on Tuesday, 22th, from a few minutes before 8 am up to almost 10 am, it was the only transformer supporting power demand. This might have been due to maintenance, or an operational issue. Either way, careful should be taken in order to separate these periods out from the analysis of the normal characteristics of the substation, and any other under investigation.

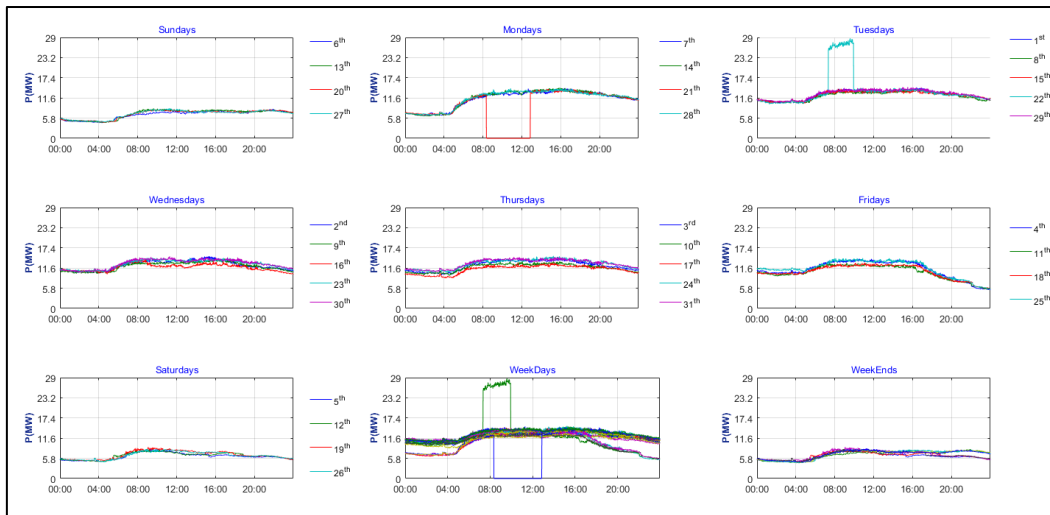


Fig. 60 – Daily Profiles registered in July/14 from one of the two parallel transformers in Kitt Green’s substation: it is possible to observe hours where the transformer is out of service (Monday, 21st), and also the hours where it was the only transformer operating the demand that was supposed to be fed by the two parallel transformers (Tuesday, 22th).

It is also possible to identify the hours of the day where the load demand is increasing, decreasing, static, low, medium or high, whether by visual inspection of Fig. 61, or by the calculation of the slopes of the daily profile curve over 1-hour successive windows. Fig. 62 analyzes a residential daily profile and it is possible to cluster the hours of the day that the demand is increasing, decreasing, and static.

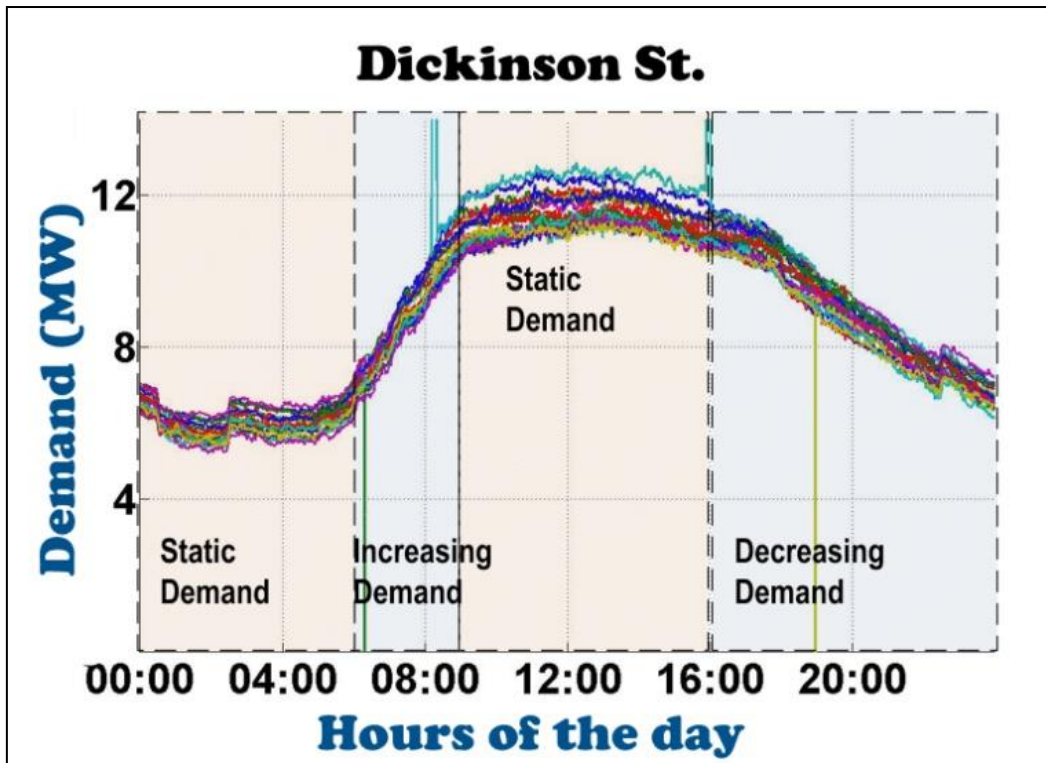


Fig. 61 – Diurnal patterns of load demand for a bell-shaped daily profile

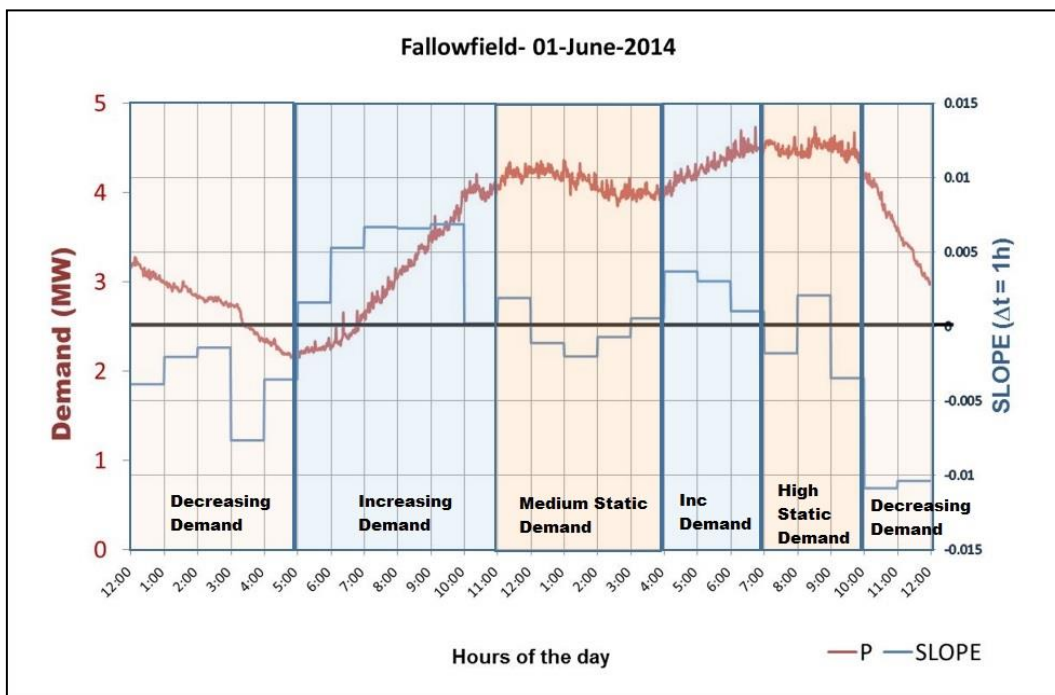


Fig. 62 - Diurnal patterns of load demand for a residential daily profile

To perform a statistical analysis of large amounts of data, it is very important to make an effort in this preliminary pattern recognition, in order to separate the most useful groups of data for the load modeling and CVR factors calculation.

5.2. Data treatment

To estimate load models that translate CVR effectiveness, it was designed computational procedures to separate and select suitable data, aided by exploratory data analysis and other statistical tools. It was investigated filtering techniques to smooth voltage and power data, in order to mitigate noise effects, whether caused by random (or spontaneous) load variations, or due to noise measurements; spontaneous load variations influence that can be of the same order (or even bigger) of the natural response of the load to small voltage variations.

It is worthwhile to highlight the concept of the spontaneous load variations that are not correlated to voltage variations. During the day, countless end-user devices are connected and disconnected all the time, leading to spontaneous or natural load variation. This is the main source of uncertainty in the assessment of load sensibility to voltage variations. In this context, noise effects mean any uncorrelated signal “added” or “subtracted” to the measurement device records, due to transducers inaccuracy that introduces small variations to the resulting equipment readings and recordings.

The recorded data from the primary substations were extracted from text files and stored into tables of a relational database. Each substation was analyzed for the whole year period, but to identify the voltage steps performed by transformer tap changes, it was necessary to develop an automatic procedure to obtain exactly:

- The instant t_0 immediately before the voltage step.
- The instant t immediately after the voltage steps and whether it is a voltage reduction or a voltage increase.
- If, after a certain amount of time, any more voltage steps occur, and if it is in the same direction as the previous step or not.

The next subsection will briefly describe the voltage steps identification procedure, in order to calculate CVR factors using precisely the time instants before and after the voltage control action.

5.2.1. Voltage Steps Identification Procedure

The solution to identify the voltage steps was the calculation of the slope of the voltage, through a sliding window, as it is illustrated in Fig. 63. The slope is the vertical distance divided by the horizontal distance between a set of points, and it can be interpreted as the rate of change in voltage (or power, as it was shown in Fig. 63) as time move forward. Its equation is formulated in (5.2):

$$slope_i = \frac{\sum_{i}^{i+\Delta t} \left[t_i - \left(\sum_{i}^{i+\Delta t} t_i / \Delta t \right) \right] \left[V(t_i) - \left(\sum_{i}^{i+\Delta t} V(t_i) / \Delta t \right) \right]}{\sum_{i}^{i+\Delta t} \left[t_i - \left(\sum_{i}^{i+\Delta t} t_i / \Delta t \right) \right]^2} \quad (5.2)$$

Where:

- t_i is the i^{th} sample of time.
- Δt is the size of the sliding window.
- $V(t_i)$ is the i^{th} sample of voltage, recorded at instant t_i .

The voltage steps identification procedure can be illustrated in Fig. 64, showing, in the upper plot, the voltage in p.u. (red line series and red axes) recorded during a CVR trial, and the slope calculated through a 10-second sliding window (green line series and green axes). The slope calculated through the 10-second sliding window is negative when the voltage tends to decrease within this 10-second window, and is positive when the voltage tends to increase. The active power data and the reactive power data are also presented, as well as their calculated slopes, confirming graphically the hypothesis that a voltage decrease causes a demand reduction, and that a voltage increase causes a demand boost.

The slope time-series detects the voltage steps whenever the value calculated with (5.1) exceeds a certain bound. The ± 40 in the right axes of the upper plot of Fig. 63 are the empirical limits determined in this survey, which indicates a voltage variation, due to tap change operation, scheduled within CLASS trials, defined in Fig. 53.

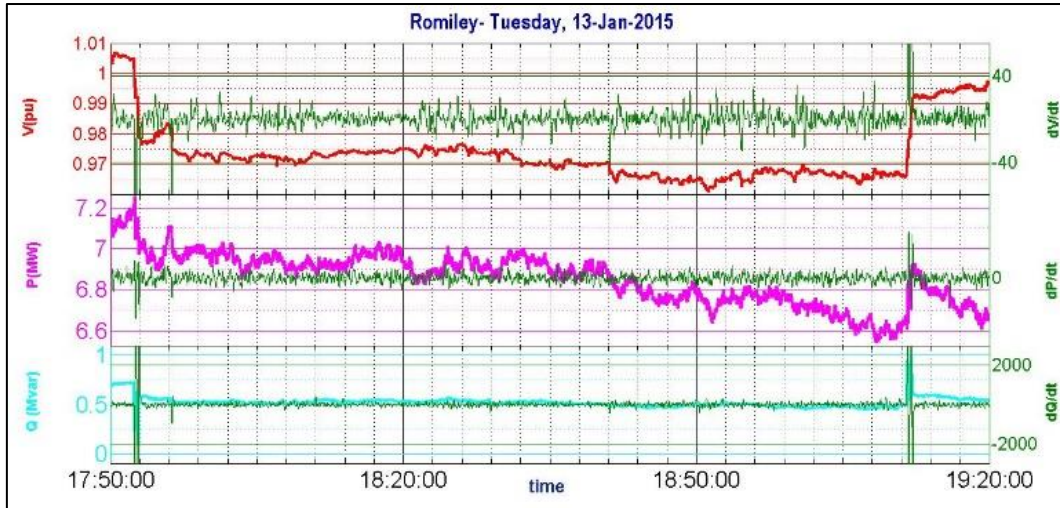


Fig. 63 – Slope calculation (green curves) to identify voltage steps

It can be observed in Fig. 63 that the first voltage step was detected at $t_0=17:52:18$ (slope magnitude greater than 40), reducing the voltage from 1.005 p.u. to 0.992 p.u. and a few seconds later, another voltage step, in the same direction (voltage decrease), reducing from 0.992 to 0.979. At 17:56:08 it was detected another voltage reduction, from 0.982 p.u. to 0.975 p.u. At 18:41:01 there is another small voltage reduction, and finally at 19:11:52 the voltage is increased from 0.979 p.u. to 0.992 p.u. In this example, to calculate Kp , as formulated in (2.10), the instants t_0 and t that correspond to the pre-disturbance and post disturbance instants are respectively 17:52:18 and 17:56:08. The pre-disturbances values of voltage, active and reactive powers are calculated as the averages within one minute (from 17:51:18 to 17:52:18).

The criteria for selecting **events** for parameter estimation of CVR factors in this work is listed below:

- Each voltage step reduction provided by the OLTC was named one-directional step (1 dS), every aggregation of more than one OLTC tap reduction was named one-directional multi-step (1 dMS). This indicator identifies a Trial 1 example, as defined in the table of Fig. 53.
- It was only selected for the estimation voltage reductions (1dMS) greater than 2%.

From each event selected for CVR factors estimation, it was retrieved from the database 2 minutes before the first voltage tap change reduction, and 8 to 10 minutes after the last voltage tap change reduction.

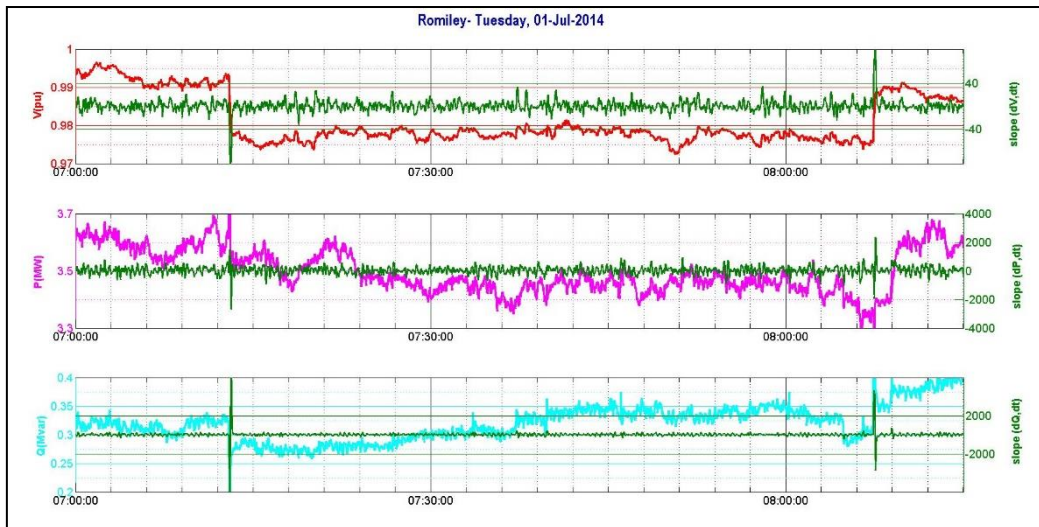


Fig. 64 - Example of the detection of a two-directional multi-step (OLTC successive tap changes reducing voltage a after a some time the OLTC successive tap changes boosting voltage)

It may be clear at this point that the correct determination of these instants t_0 and t are determinant for CVR factors assessment, because different windows sizes may produce different voltage/power post-disturbances values, due to power deviations not related to voltage deviations, defined as natural load variations.

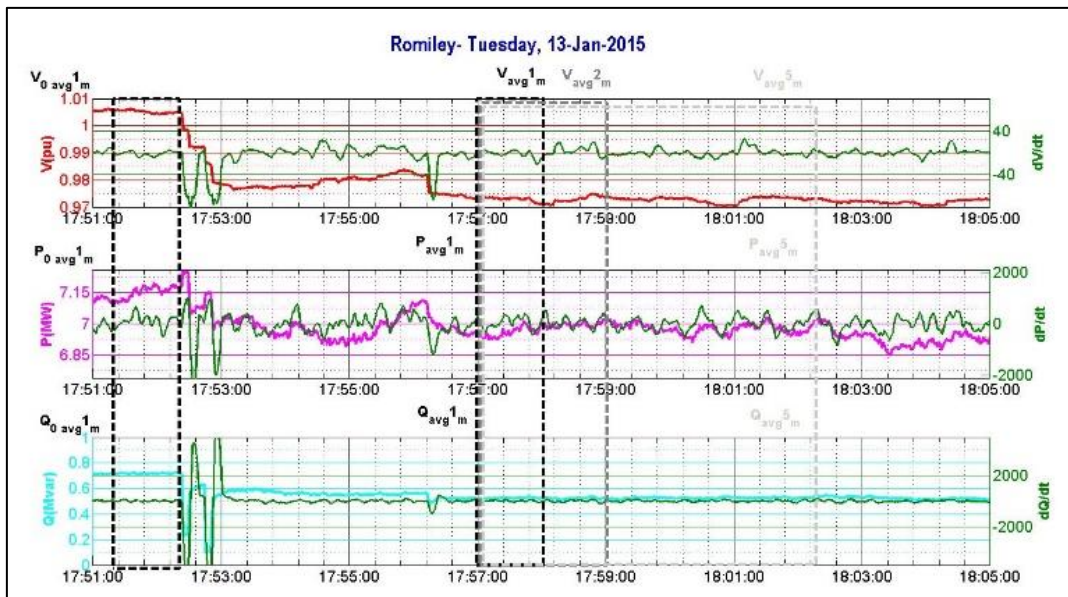


Fig. 65 - The choice of the window size is not trivial

Both demand and voltage signals were subjected to a smoothing filter to eliminate noise and missing data. Both signals were normalized using the data immediately before the first voltage reduction occurred.

Savitzky-Golay is a digital filter for smoothing the data, increasing the signal-to-noise ratio without significant distortion of the signal. Savitzky-Golay smoothing

filters (also called digital smoothing polynomial filters or least-squares smoothing filters) are typically used to "smooth out" a noisy signal whose frequency span (without noise) is large [57]. Fig. 66 presents the implementation of the Savitzky-Golay filter using Matlab built-in function, processing this 14-minute P and V measurements from one of the transformers installed in Romiley.

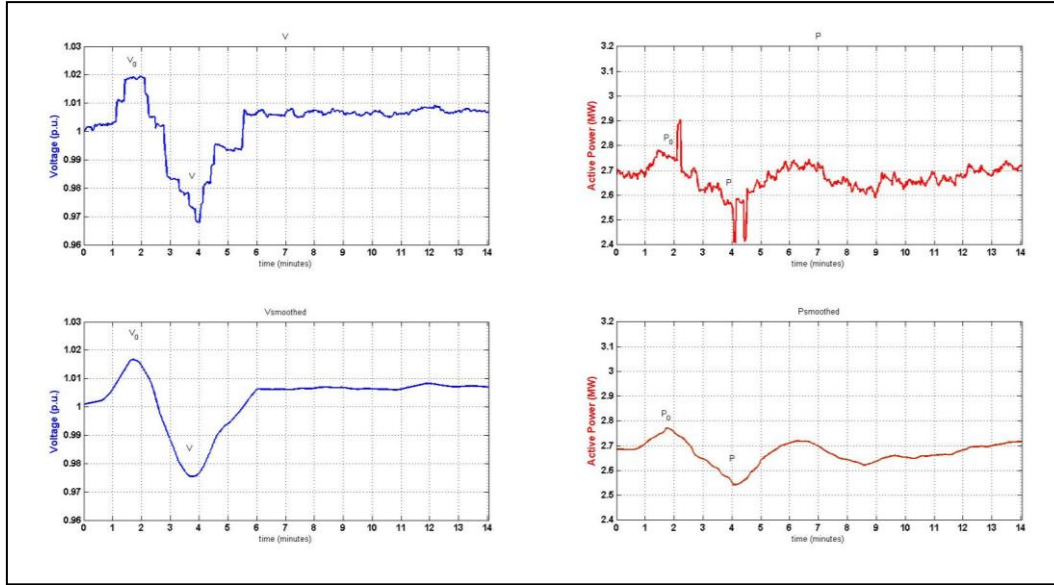


Fig. 66 – Voltage and active power data (upper plots) and their Savitzky-Golay smoothed counterparts below.

5.3. Model Selection

In section 2.2.1.5, it was formulated a nonlinear dynamic load model with 3 parameters, a transitory Kp (which will be α_t , in (5.4)), a time constant (which will be T_p , in (5.4)) that counts, in seconds, the partial load recovery, and a steady-state Kp (which will be α_{ss} , in (5.4)), that can be interpreted as the actual effect that a voltage step produces in the load.

In [54] CVR factors are defined in the same way as the parameters Kp and Kq of the traditional exponential load model, reviewed in section 2.2.1.5, formulated in (2.8). Thus:

$$P(t) = P_0 \left(\frac{V(t)}{V_0} \right)^{Kp} = P_0 \left(\frac{V(t)}{V_0} \right)^{CVR_p} \quad (5.3)$$

In the next section, the parameter estimation results will use the Exponential Recovery Model, formulated again in (5.4), in order to quantify CVR effectiveness, and the variations expected in this parameters through diurnal or annual seasons.

$$T_p \frac{dP_r}{dt} + P_r = P_0 \left(\frac{V(t)}{V_0} \right)^{\alpha_{ss}} - P_0 \left(\frac{V(t)}{V_0} \right)^{\alpha_t}$$

$$P_l = P_r + P_0 \left(\frac{V(t)}{V_0} \right)^{\alpha_t}$$

(5.4)

Parameters set: $\theta_p = [\alpha_{ss}, \alpha_t, T_p]$

Subjected to:

$$\begin{aligned} \alpha_{ss} &< \alpha_t \\ \alpha_{ss} &\geq 0 \\ \alpha_t &\geq 0 \\ T_p &> 0 \end{aligned}$$

This *a priori* model selection [21]-[22] is justified because peak demand reduction and energy saving are interested in evaluating a steady-state load response, and the parameter α_{ss} is a better estimation of the CVR effectiveness than α_t , because as it is reported in the results contained in [2] and [3], the time constant is no more than a few minutes, and then the load restores to a new level, that is not so severe as the transient reduction, but it is usually below the pre-voltage reduction level. To illustrate this, the load model parameters of the model structure (5.4) in the example of the Fig. 67 are $T_p=10$ seconds, which is the amount of time between 30 to 40 seconds; $\alpha_t=2$, which indicates that for a voltage reduction of $\Delta V=0.1$ p.u., there is an ideal transitory demand reduction of $\Delta P=0.2$ p.u.; and $\alpha_{ss}=1$, which indicates that for $\Delta V=0.1$ p.u., there is an ideal steady-state demand reduction of $\Delta P=0.1$ p.u.

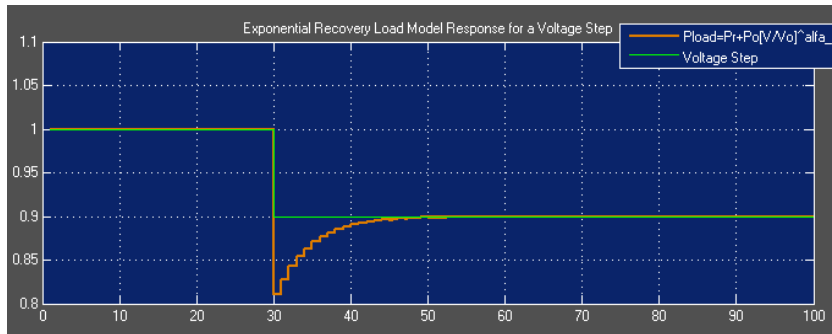


Fig. 67 – Step Response of the Exponential Recovery Load Model (5.4)

In section 5.2.1, it was presented a detailed description of the voltage step identification procedure, highlighting the importance of determining the exact instants t_0 and t to calculate the voltage demand sensibility, which is the core of this load modeling investigation. In order to assess CVR factor using this load model formulation, the post-disturbance starting instant t cannot be too close to the end of the voltage step, because there is a predictable partial load recovery, neither it can be too away from the end of the voltage step, because of the also predictable load spontaneous variations. In [3], it is suggested that the time between the initial of voltage step and the post-disturbance to use for analysis could be around 2.5 to 3 times de time constant Tp .

In the following sections, it will be presented some results from the parameter estimation using data collected from the substations of CLASS project. Each event was estimated separately using the Simulink Design Optimization of Matlab. The optimization method for parameter estimation was also a Genetic Algorithm, using MatLab's optimization toolbox.

5.4. Parameter Estimation

Equation (5.5) extends (3.2), because the sum of squared errors is weighted proportionally to the voltage deviation from 1 p.u.. Thus, when the optimization routine is searching for a good parameter set that is able to model load response, the period(s) of the greatest voltage deviations are the most important to be modeled accurately. Therefore, the chosen objective function to be minimized is the weighted sum of square errors (SSE), but instead of setting the weights inversely proportional to the standard deviation of the power demand, they were set proportional to a kind of normalized voltage deviation from 1 p.u..

$$F_{obj}(\theta_p) = \frac{1}{n} \sum_{t=1}^n \omega_{pt} \left(\frac{P_{meas}(t) - P_{calc}(t, \theta)}{P_{meas}(t, \theta)} \right)^2 \quad (5.5)$$

$$where: \omega_{pt} = |1 - V_{meas}(t)| / \sum_{t=1}^n |1 - V_{meas}(t)|$$

Where:

- F_{obj} – the objective function to be minimized.
- θ_p is the active load model parameters set.

- n – the total of samples recorded during an event.
- t – the variable that indexes the order of the samples.
- $P_{meas}(t)$ is the three-phase active power data measured, with n samples.
- $V_{meas}(t)$ is the one voltage to neutral phase data measured, with n samples.
- $P_{calc}(t, \theta_p)$ is the three-phase active power data calculated by the chosen load model, with n samples, and parameterized by θ_p .

After selecting a set of events characterized by voltage reductions greater than 2 %, the data is filtered aided by the Savitzky-Golay filter, and then the parameter estimation is performed in the Simulink Design Optimization package, which provides functions, interactive tools, and blocks for analyzing and tuning model parameters. Fig. 68 and Fig. 69 presents data from the active and reactive power demand respectively, and also data from the voltage perturbation, all of them normalized by their pre-disturbance values, recorded in the same event. The data shown in these pictures are not filtered, it is the raw data, where the blue curves are the measured active and reactive power and the magenta curves are the measured voltage to neutral data. The red curves are the active and reactive power data simulated by (5.4), with the estimated parameters registered in the top of the plots.

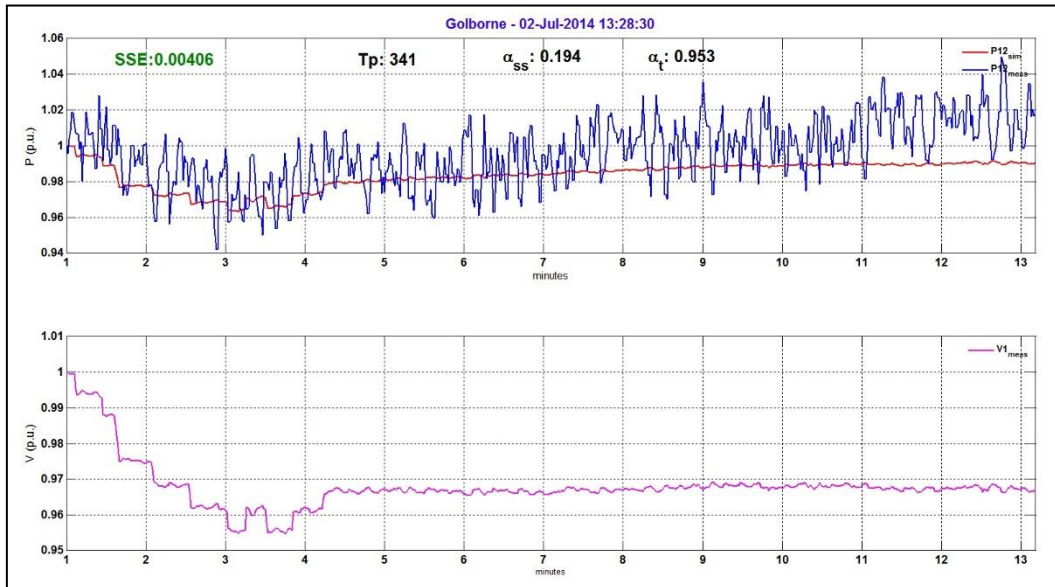


Fig. 68 – Active power load model parameters estimation for a summer event for the Golborne substation.

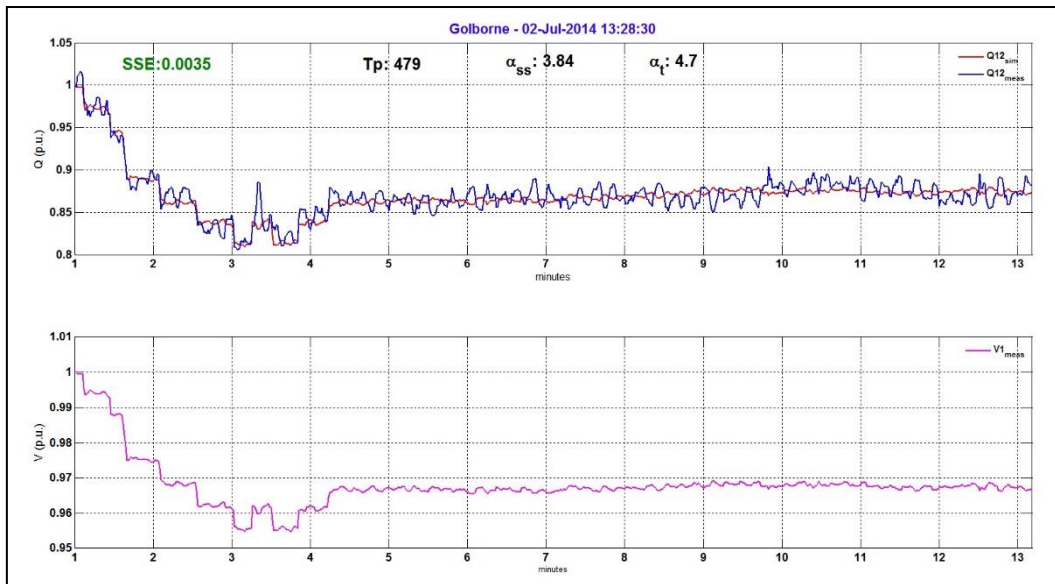


Fig. 69 – Reactive power load model parameters estimation for a summer event for the Golborne substation.

The parameters estimated for active power are much smaller than the ones for reactive power, and this was expected since the voltage has a stronger influence in the reactive power flows, and also the latter is a much more nonlinear phenomenon.

Due to the large amount of data collected, this parameter estimation approach used the concepts of statistical inference to treat the collection of estimates of CVR factors, obtained for each voltage reduction event. The next subsection describes some of the concepts used to derive CVR factors.

5.4.1. Statistical Inference

Data analysis is a wide area that can extract information and knowledge from data observations whose properties may be correlated. Statistical Inference is a data analysis approach that seeks to describe properties from a collection of data, whether by testing hypothesis or estimating a quantity of interest. It is based on the characteristics obtained from a subset of the whole data set. The aim of statistical inference data analysis is about to derive properties of an observed (measured) data set, assuming the data assumes a pattern regarding its variability over a certain range of probable values. Therefore, the statistical inference conclusions are valid under certain assumptions, such as data underly probability distribution.

In this context, the whole data set will be called *population* or *universe*, and the data subset, a *sample* of the *population*. Numerical descriptions of a *population*, like *proportions*, *averages* or *standard deviations* are called *parameters* of the

population, whereas the same indicators for *samples* are called *statistics*. Statistical inference provides tools to estimate *parameters* from *statistics*, leading to statistical conclusions from samples, supported by probability theory.

The difference between the concepts of *point estimation* and *interval estimation* is straightforward: a *point estimate* is an estimate of a single population parameter such as, for example, the mean of a quantity of interest. *Interval estimation* is a range of values that the same quantity of interest most likely lies within, under a predefined probability (*confidence level*).

Thus, the main goal is to use statistical inference for CVR assessment within a *confidence interval* [59]-[61].

Equation (5.5) estimates the expected value for α_{ss} within the entire *population*, from a *sample*, within a range with 95% of confidence:

$$\hat{x}(P \leq 95\%) = \bar{x} \pm 1.96 \frac{\sigma_s}{\sqrt{n}} \quad (5.6)$$

$$\sigma_s = \sqrt{\frac{(x - \bar{x})^2}{n-1}}$$

Where:

- \hat{x} – is the expected value for the *parameter of the population*.
- \bar{x} – is the average of the *statistics*, i.e. the average of x calculated from the *sample*.
- σ_s/\sqrt{n} is the standard error of the expected value \bar{x} , n is the size of the *sample*.
- σ_s is the *sample standard deviation*, used in place of the *population standard deviation*, which is unknown.

The factor 1.96 is valid only if:

- the distribution of the sample means of x follow a normal distribution,
- the intended confidence interval is 95%,
- the sample size is greater than 30.

If the confidence interval target is 99%, 1.96 should be substituted by 2.58, according to the Gaussian distribution tables [60]. If the sample size is less than 30, this number should follow a *Student's t distribution* [61].

A *sampling design* or *scheme* must ensure that a representative *sample* is chosen. If the *population* is heterogeneous as a whole but can be split into

homogenous sub-populations, the *stratified sampling* is supposed to provide better results than a *simple random sampling* within the whole population, or to reduce variance of the estimates within the sub-populations, which is also known as *clusters*.

One important assumption to be done is that the parameter distribution should follow a normal distribution, spread over its expected value. After the stratification of the *population*, divided in a way that the groups or *strata* samples present similar characteristics within the *stratum* and then a *simple random sample* should extract a *sample* containing data from each *stratum*, proportionally to the size of each *stratum* within the *population*.

Data collected through an entire year period can be divided taking into account seasonality, days of the week (weekdays, Saturdays, Sundays and holidays) and the hours of the day (as shown in Fig. 61 and Fig. 62). Thus, the next list divides the data accordingly:

Season:

S1 – from June to August 2014.

S2 – from September to October 2014.

S3 – from November 2014 to February 2015.

S4 – from March to May 2015.

Days of week:

W1 – From Monday to Friday.

W2 – Saturdays, Sundays and Holidays.

Day hours:

H1 – [22hs to 5hs) - Low Static/Decreasing Demand.

H2 – [5hs-10hs) - Increasing Demand.

H3 – [10hs-16hs) - High Static Demand (commercial/industrial) or Medium Static Demand (residencial).

H4 – [16hs-19hs) - Decreasing Demand (commercial/industrial) or Increasing Demand (residencial).

H5 – [19hs-22hs) - Decreasing Demand (commercial/industrial) or High Static Demand (residencial).

Table 4 - Group definitions for strata specification

Groups description	code	from	to	N	n
Summer ~ Jun-Aug/2014	S1	2014-06-01	2014-08-31	92	25.2%
Autumm ~ Sep-Nov/2014	S2	2014-09-01	2014-11-30	91	24.9%
Winter ~ Dec-Feb/2015	S3	2014-12-01	2015-02-28	90	24.7%
Spring ~ Mar-May/2015	S4	2015-03-01	2015-05-31	92	25.2%
Low Static/Decreasing Demand	H1	22:00	05:00	7	29.2%
Increasing Demand	H2	05:00	10:00	5	20.8%
High/Medium Static Demand	H3	10:00	16:00	6	25.0%
Decreasing/Increasing Demand	H4	16:00	19:00	3	12.5%
Decreasing/High Static Demand	H5	19:00	22:00	3	12.5%
Weekends	W2	Saturday	Sunday	2	28.6%
Weekdays	W1	Monday	Friday	5	71.4%

The stratification may be used to focus on a particular subset of the available data sample, or to combine statistics from the different groups taking into account their proportions within the population. For example, in this case study, Table 4 leads to the obvious conclusion that weekdays represent the majority of the days in a year, and the data sampled from weekdays must be the most important to be analyzed. It should be more relevant to sample from weekdays, than from weekends or holydays, unless the purpose of an investigation is precisely to understand whether CVR action is valuable to be applied in weekends or not.

For peak demand reduction, Fig. 54 to Fig. 61 has shown that the peak hours (H3) of weekdays (W1) are likely to be the scenarios where the highest demand will occur. In the wintertime (S3), the heating devices in Manchester, UK represent a big share of total power demand.

To obtain the high static demand H3, it was designed a query to the database to aggregate the peak hours of each weekday of the year in a graphical way, following the EDA approach to support this data stratification. The histograms are the perfect tool for that, and the next figures will show the peak demand hours by substations of the three types. Fig. 70 presents the peak demand hours distribution, where the first column of graphics shows the substations classified as Type 1 (Avenham, Kingsway, Blackpool), the second column shows the substations Type 2 (Fallowfield, Romiley, Golborne) and the third shows the Type 3 (Anne Pit, Kitt Green, Buckshaw).

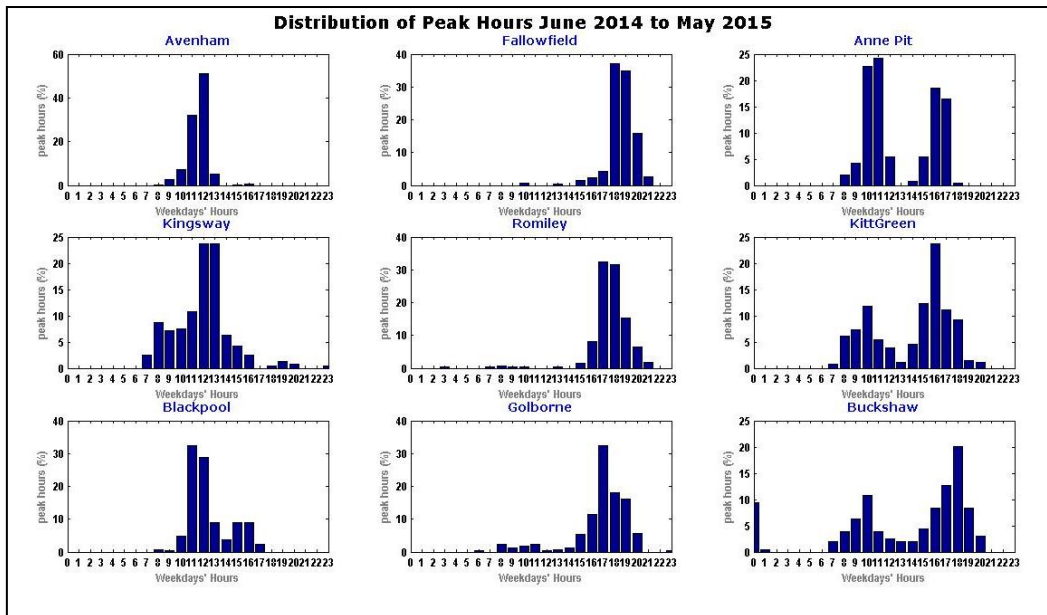


Fig. 70 - Histograms of 9 substations analyzed in the CLASS project.

It is interesting to note that Type 1 and Type 2 registered the peak hours respectively around 11am/12am and 6pm/7pm, and the Type 3 ones behave with characteristics from both Type 1 and 2. Anyway, the parameter estimation will begin to analyze data from these hours of the day, for each one of them, since it is the period where this ancillary service would be most welcome.

5.4.1.1. Bootstrap method

To infer the characteristic values of a population, and also the margin of error of such inference, it is very common to assume that data follows a distribution pattern. The most used assumption is that the underlying distribution of data follows a Gaussian curve that can be fully determined by two parameters (mean and variance). There are a few tests to check if this is likely to be true or not. Whenever it is not possible to identify if data follows any known statistical distributions, the bootstrap method can be used.

The bootstrap method is an application of Monte Carlo method, and it is a sampling strategy based on sampling with replacement of the data units' multiple times (typically, one thousand, two thousand times). The resulting simulated sampling distribution is used to calculate an interval of plausible values that the unknown parameter is likely to lie within.

Fig. 71 shows 9 normal probability plots that compare the distribution of the CVR factors in the x-axis to a normal distribution in the y-axis. Each CVR factor is

plotted using the plus sign ('+') markers that should follow a reference line that represent the theoretical distribution. A solid reference line is drawn between the first and third quartiles of the CVR factors' distribution, and a dashed reference line extends the solid line to the ends of the data. If the CVR factor's distribution follows a normal distribution, then the data points appear along the reference line. On the other hand, if the distribution of CVR samples reveals a different distribution pattern, a curvature can be seen in this plot. Thus, these normal probability plots reveals graphically if one can assume whether a distribution follows a normal one.

The results presented in Fig. 71 indicate that a sampling design based on the assumption that the data has a normal distribution may introduce bias and therefore the bootstrap method was used.

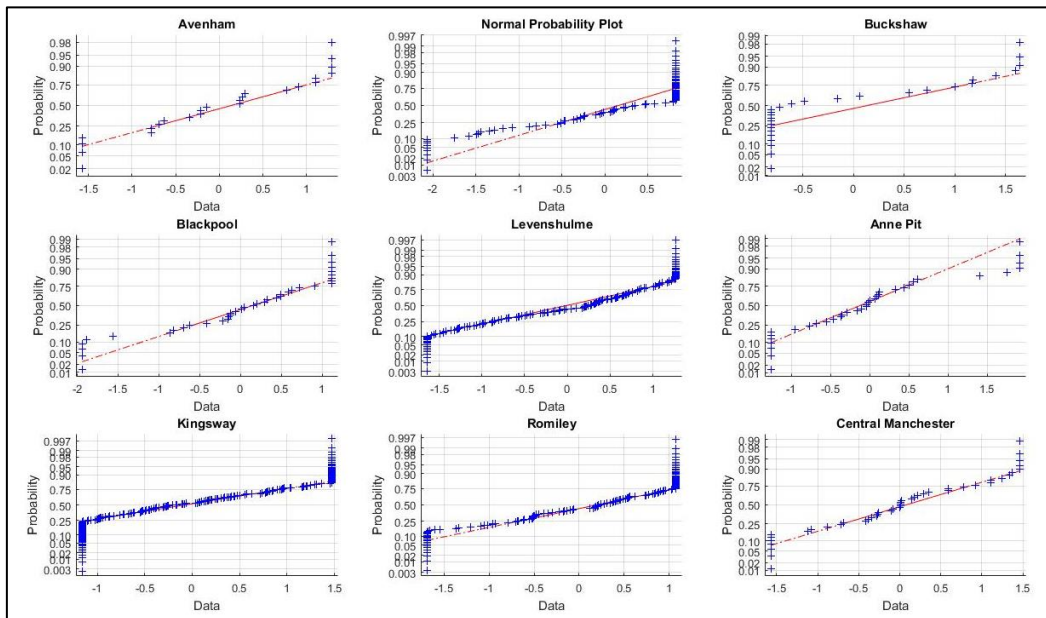


Fig. 71 – Normal probability plots one the substations of Type 1, Type 2 and Type 3, presented at the first, second and third columns respectively.

Fig. 72 presents the impact of bootstrap methods on time series analysis in a graphical way. There were 72 samples available for the substation Anne Pitt, and the histogram at the bottom of the first column is clearly not following a Gaussian curve. However, after the bootstrap resampling method, the mean of the means distribution has good adherence to the Gaussian Curve, and the histogram at the bottom of the second column appears to be much similar to the expected shape of a Gaussian curve.

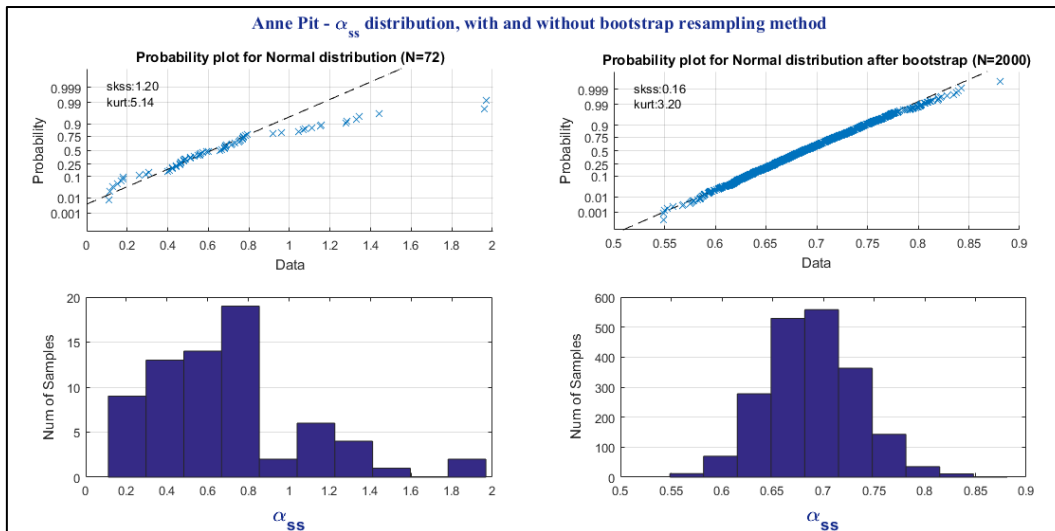


Fig. 72 – An EDA approach to justify the bootstrap method to infer the expected value of CVR factor α_{ss} : the first column presents the normal probability plot of the α_{ss} distribution as it was estimated and the histogram of the same data distribution; the second column presents the same plot after the bootstrap resampling method.

5.5. Results

Table 5 presents the results from 10 substations chosen to illustrate the results using the methodology described up to now in this thesis. This table includes the weekdays' average energy (MWh), the CVR factors and their respective 95% confidence intervals calculated with the bootstrap method. In order to provide a better visualization of the results, the steady state CVR factors are shown in Fig. 73, as well as the other estimated parameters of (3). These graphics are known as boxplots, wherein each box, a central mark indicates the data median, or the 50th quartile of data distribution. The box boundaries are the 25th and 75th quartiles. The dotted lines, or “whiskers”, indicates data up to the 25th quartile, and above the 75th quartiles. Quartiles divide ordered data distribution (from smallest to largest values of the data distribution) in four parts (the first part is the 25% smallest data, and the last part represents the 25% greatest data of the sample distribution under study).

Table 5 - CVR Factors expected values and their confidence intervals

Substations	Classification		MWh (daily avg)	CVR _{avg} conf. interval	CVR _{avg}	CVR _{avg} conf. interval	Sample Size
SSML1	Mixed	NR	287	0.58	0.67	0.76	68
SSML2			228	0.63	0.72	0.80	121
SSML3			191	0.97	1.05	1.14	146
SSIC1	Ind/Com	NR	158	1.18	1.36	1.54	59
SSIC2			129	1.06	1.25	1.45	46
SSR1	Residential	R	398	0.73	0.82	0.91	121
SSR2			224	0.87	0.98	1.08	111
SSR3			209	0.93	1.02	1.11	164
SSR4			164	1.07	1.19	1.32	123
SSR5			174	1.33	1.48	1.63	92

In Fig. 73 the substations are in the same order, from left to right, as they are listed in Table I, grouped by their type: mixed, mainly industrial/commercial and mainly residential. The substations grouped by types are ordered from the most loaded to the least loaded. The top plot in the Fig. 73 shows that all of the CVR factors' distributions have higher medians in the substations less loaded. The last boxplot in every graphic is the parameter data distribution from all substations, and it presents what it can be expected, in general, from CVR factors.

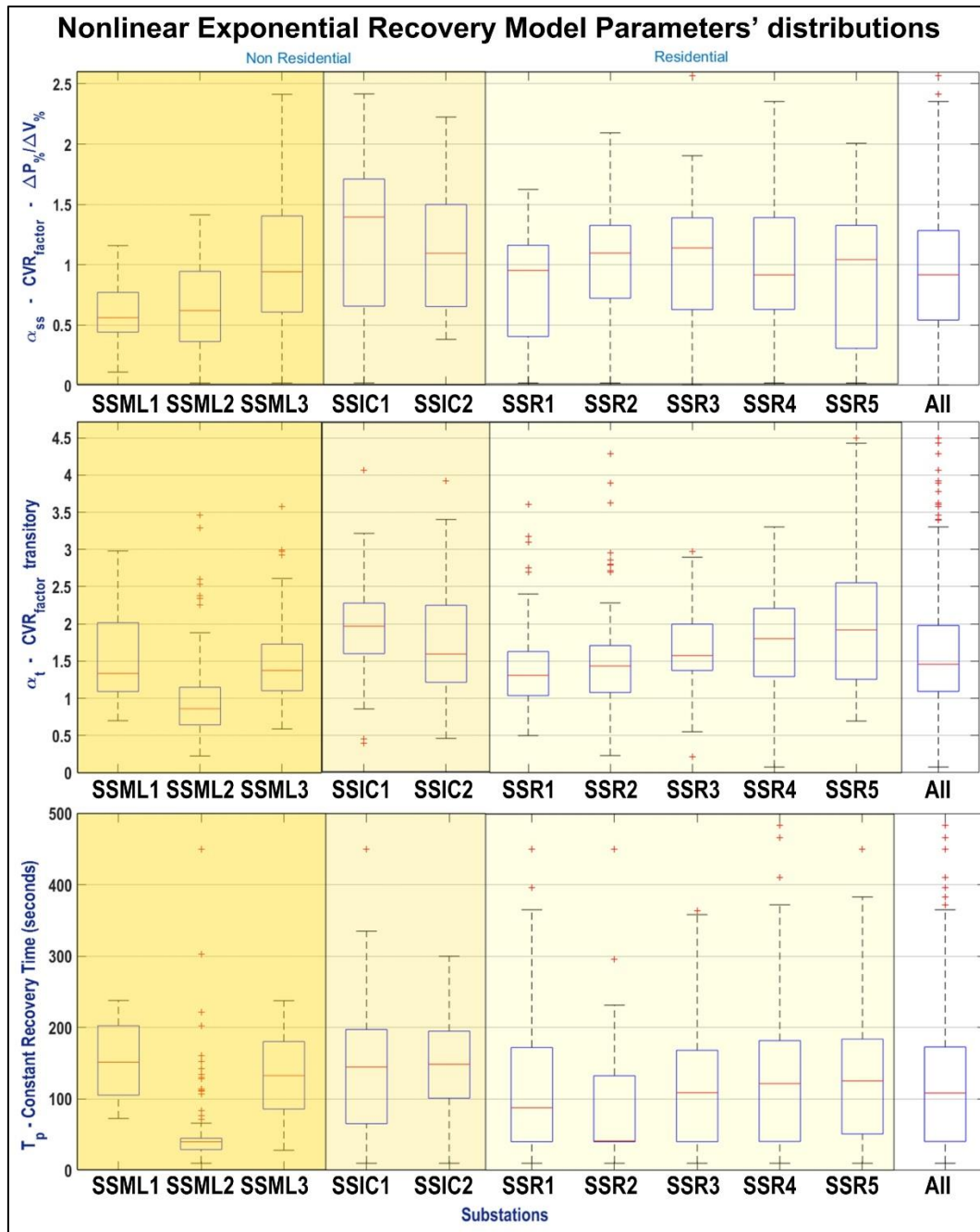


Fig. 73 - CVR factors estimates' distributions

Fig. 74 investigates if there is a correlation between the loading conditions and the effectiveness of the CVR to reduce peak demand. The trend line plotted, and the corresponding equation in the figure, can be understood as a weak correlation between loading conditions and CVR factor, and since the equation of this regression line contains a negative slope, it indicates that the more loaded the system is, the less effective the peak demand reduction will be. Eight out of ten substations, whose correlations have been investigated individually presented a

similar pattern, i.e., a weak negative correlation, which confirms the above conclusion.

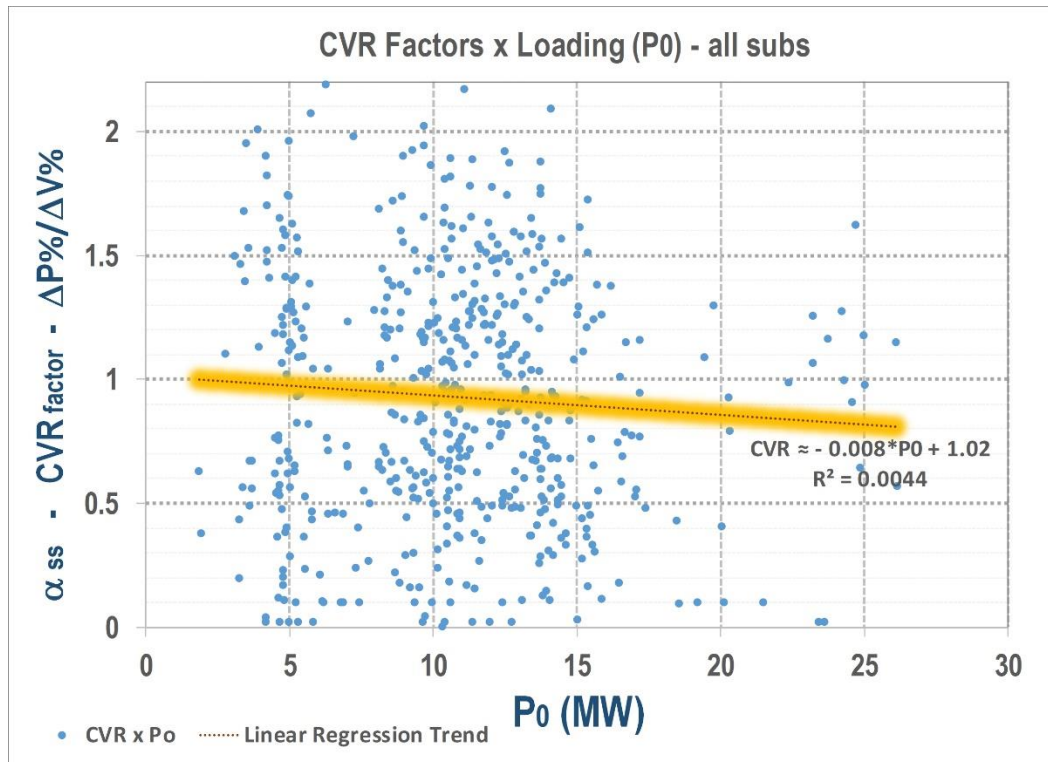


Fig. 74 - Scatter plot correlating CVR factor against loading conditions

5.6. Summary of the Chapter

The CVR assessment is an application of the load modeling methodology, because:

- It was solved using a load model selection made *a priori*, which means that this mathematical structure was not compared to others, since the phenomenon is characterized by a temporary load response greater than the final or steady state load response, and the exponential recovery load model's structure contemplates this feature; furthermore, the graphical results of these parameters confirmed this theoretical knowledge.
- The parameters set was estimated, for each event of voltage reduction performed by the OLTC transformers, by means of an optimization procedure based on Genetic Algorithms.

Because the data were obtained for many different scenarios, it was demanded statistical tools to cope with the uncertainty underneath the conclusions and the heterogeneity of the data distribution.

The CVR evaluation is dependent on the load mix and the loading conditions that are correlated to weather conditions, the hours of the day and to season characteristics. Nevertheless, the available data from the substations investigated in this work could not identify any pattern when correlating CVR factors to season changes.

Demand daily profiles also demonstrate visually the loading patterns that repeat, day after day, and the slope calculated over a 1-hour sliding window size aided to cluster the identified patterns into increasing demand, decreasing demand and static demand, where the latter is the best period of the day to calculate CVR factors, due to the small slope.

The CVR factor data distribution of Fig. 73 shows that this indicator is within 0.5 and 1.5, which means that every 1% voltage reduction is capable of achieving between 0.5% and 1.5% of demand reduction.

The nonlinear exponential recovery load model was chosen because the parameter estimation procedure evaluates the temporary and steady state demand response for each voltage reduction event. This distinction was essential to avoid overestimation of the effectiveness of CVR approach, since the instantaneous demand response (transient CVR factor α_t , the middle plot of the Fig. 73) is typically greater than the steady state demand response (steady state CVR factor α_{ss} , the upper plot of the Fig. 73). Furthermore, the statistical data analysis concluded that the expected load response time from transient to steady state is typically from 1 up to 3 minutes (Constant Recovery Time T_p , bottom plot of the Fig. 73).

The correlation analysis plotted in Fig. 74 indicates that CVR is less effective as load demand is higher, despite of the weak correlation.

6 Conclusions

The methodology proposed in this thesis intended to describe a set of procedures, inspired by system identification concepts, which can be implemented using many different techniques and many available software resources, but there are some macro aspects that must be included in the systematic modeling procedures that were discussed theoretically and practically in this thesis: sampling design, data processing, parameter estimation and some form of model validation are the core processes, but definitely there is not an unique form to implement each one of these tasks. References [2] and [63] describes the most used load models and a few techniques frequently used to estimate load model parameters.

It is recognized that load modeling is a key aspect for power system stability, therefore it is necessary to continuously update and validate the way loads are represented in power systems simulations. It has been reported in the literature that static load models, reviewed in section 2.2.1, may not be able to represent load behavior during or after voltage disturbances, particularly for large voltage variations. Field measurements of reactive power behavior during voltage steps below 0.9 p.u. showed that the flux of reactive power can temporarily change its signal sign, which a static ZIP model is mathematically unable to represent.

It has been shown along this thesis that loads are generally voltage-dependent, and thus it influences the stability margins, because representing loads in a simulation that overestimate a voltage disturbance may result in a very conservative security margin evaluation, leading to a poor utilization of transmission lines. Conversely, load models that underestimate voltage disturbance may lead the system to operational points too close to the limit of stability, and a voltage collapse can be the worst consequence of this poor design.

There are many parameter estimation techniques and this choice is a matter of expertise of the developer, or associated to the availability of these techniques built in software development environments. The methodology was described in a way that it does not really matter what type of model is chosen, nor the parameter

estimation technique. Most optimization techniques have their pros and cons, and one needs an extra effort to configure properly the method options, and prepare data to search the space solutions efficiently. Nevertheless, optimization problem that seeks to minimize or maximize a cost/error functions always may stuck in local minimum solutions, and sometimes small tricks in the implementation may prevent this, such as the investigation of different initial parameter guess to start the optimization, that can be obtained from previous knowledge of experts, or from literature, although for aggregated load is not really easy to find universal parameters that are supposed to be always accurate and representative. The objective function definition can also be done in a creative way, and the results may be combined with statistical tools seeking to extract knowledge from historical measurements database.

During the case study presented in chapter 4, the optimization procedure was built in a way that parameter estimation could seek for a parameters' set that tries to respond to a range of voltage variations that were likely to occur during a year or two. The reactive power results showed that constant impedance load model, used in the Brazilian interconnected system representation of 69kV systems, is frequently underestimates the real deviations in the reactive power after voltage sags. It is worthwhile to recognize that reactive power load modeling is the most difficult task to achieve, because there are reactive compensation devices downstream the measuring device placement that will change the reactive flows of the response of the aggregated load, but it is hard to find out whether these reactive compensation devices are on or off during the different voltage sags registered.

In chapter 4 it was given details of data processing, because the data measurement devices raw data is recorded with different sampling rates that determine the precision to filter data into the harmonic components of interest. The discrete Fourier transform was able to extract from instantaneous voltages and currents, the fundamental components that evaluated active and reactive power. The measurement device must be installed using the load convention, therefore wrong terminal connection and the identification of the phase sequence (ABC or ACB) will affect the results of the signal processing.

In the case study of chapter 5, it was investigated the peak demand reduction through the voltage demand sensibility, using the load modeling methodology described in chapter 3. The field-data measurements were obtained from a

conservation voltage reduction (CVR) project in Manchester-UK, where one of the goals was to quantify the effectiveness of this technique and the potential to provide ancillary services to the transmission system operator, through a modern communication data exchange.

The results proved that the effectiveness of CVR is dependent on the load mix and the loading conditions. An automatic method for the estimation of load model parameters supported the CVR factors assessment, but to select the data that would be representative in a statistical way, it had to be developed a voltage step detector, and a parameter estimation procedure to fit a nonlinear model into the data selected and smoothed.

As a comparison, in the case study of the chapter 4, the time frame was much smaller, but the voltage variations were much larger. In CVR assessment, the voltage variations were up to 6%, and because the load demand varies in minutes by a few percent, it is not trivial to separate spontaneous load variations (i.e. load variations not related to voltage variations), to load variations caused by very small voltage variations.

If the objective function minimized in chapter 4 was summing many events recorded on different days, the objective function in chapter 5 takes into consideration only one event at the time. This is explained due to the different purposes of each of the case studies. In chapter 4, the load model was meant to represent 69kV systems in stability studies, using a time frame from milliseconds to a few seconds that are useful for transient stability, for large voltage variations. The purpose was modeling to represent typical scenarios or to reproduce specific events of interest, i.e. modeling for simulation. In chapter 5, the purpose was more into predicting the effectiveness of CVR, keeping in mind that it is a fuzzy limit between what it was caused by voltage control and what was originated in the connection or disconnection of downstream loads (spontaneous load variations). Therefore, it seemed like a natural choice to model the uncertainty using statistical tools that would be supposed to draw more reliable conclusions about the real potential of this volt/var control application.

EDA was used extensively for identifying patterns among the data, and the insights that this graphical approach brought led to a better understanding of the problem, before the implementation of the parameter estimation. The main goal of

the EDA visual analysis was to separate data into groups that contain distinctive characteristics particularly associated to the case study.

Statistical inference was very useful to suggest the level of uncertainty that the parameters distribution presented. It is possible to obtain estimates even for small sample sizes that may be utilized for guessing the CVR range of expected values. The estimated parameters showed that the more loaded the system is, the less effective the peak demand reduction will be.

- [1] KUNDUR, P., **Power System Stability and Control**, Electric Power Research Institute, McGraw-Hill, USA, 1994.
- [2] CIGRE Working Group C4.605, **Modeling and aggregation of loads in flexible power networks**, 2013.
- [3] NAVARRO, I., **Dynamic Power System Load, Estimation of Parameters from Operational Data**, Doctoral Dissertation, Lund University, Sweden, 2005.
- [4] IEEE Task Force on Load Representation for Dynamic Performance, **Load Representation for Dynamic Performance Analysis**. IEEE Transactions on Power Systems, Vol. 8, No. 2, pp. 472-482, May 1993.
- [5] IEEE Task Force on Load Representation for Dynamic Performance, **Bibliography on Load Models for Power Flow and Dynamic Performance Simulation**. IEEE Transactions on Power Systems, Vol. 10, No. 1, pp. 523-538, February 1995.
- [6] IEEE Task Force on Load Representation for Dynamic Performance, **Standard Load Models for Power Flow and Dynamic Performance Simulation**. IEEE Transactions on Power systems, Vol. 10, No. 3, pp. 1302-1313, August 1995.
- [7] MILANOVIĆ, J. V., YAMASHITA K., VILLANUEVA S., DJOKIĆ S., KORUNOVIĆ L., **International Industry Practice on Power System Load Modeling**, IEEE Transactions On Power Systems, Vol. 28, No. 3, August 2013
- [8] IEEE/CIGRÉ Joint Task Force on Stability Terms and Definitions, *et al.*, **Definition and Classification of Power System Stability**, IEEE Transactions on Power Systems, vol. 19, No. 2, May 2004.
- [9] HILL, DAVID J., **Nonlinear dynamic load models with recovery for voltage stability studies**, IEEE Transactions on Power Systems, vol.8, n° 1, February 1993.

- [10] MAITRA, A., *et al.*, **Using System Disturbance Measurement Data to Develop Improved Load Models**, Proceedings of the IEEE Power Systems Conference and Exposition, 2006.
- [11] OPPENHEIM, A. V., *et al.*, **Signals & Systems**, 2nd ed., Prentice-Hall Inc., 1997.
- [12] CHUANG, H.-D., WANG, J.-C., **Development of a dynamic ZIP-motor load model from on-line field measurements**, International Journal of Electrical Power and Energy Systems, Volume 19, Number 7, October 1997.
- [13] REMMU, H., JIN, M., **Composite Load Modeling via Measurement Approach**, IEEE Transactions on Power Systems, vol. 21, No. 2, May 2006.
- [14] MA, J., DONG, Z., HE, R., HILL, D. J., **Measurement-based Load Modeling Using Genetic Algorithms**, 2007 IEEE Congress on Evolutionary Computation.
- [15] ZALI, S., MILANOVIĆ, J., **Generic Model Of Active Distribution Network For Large Power System Stability Studies**, IEEE Transactions On Power Systems, Vol. 28, No. 3, August 2013.
- [16] FENG, W., XIAO-PING, Z., PING, J., **Modeling and control of the wind turbine with the direct drive permanent magnet generator integrated to power grid**, in Proc. 3rd Int. Conf. Electric Utility Deregulation and Restructuring and Power Technologies, 2008, pp. 57–60.
- [17] DEHGHANI, M. and NIKRAVESH, S. K. Y., **Nonlinear state space model identification of synchronous generators**, Elect. Power Syst. Res., vol.78, pp. 926–940, 2008.
- [18] DOVAN, T., DILLON, T.S., BERGER, C.S.; FORWARD, K.E. **A Microcomputer Based On-Line Identification Approach to Power System Dynamic Load Modelling**, IEEE Trans. PWRD-2. No. 3.pp. 529-36. Aug. 1987.
- [19] OPPENHEIM, A. V., *et al.*, **Discrete-time signal processing** Prentice Hall Inc., 1998.
- [20] GENÊT, B., **On Monitoring Methods and Load Modeling to Improve Voltage stability Assessment Efficiency**, Thesis, Université Libre de Bruxelles, Belgium, September 2009.
- [21] LJUNG L.: **System Identification — Theory For the User**, 2nd ed., PTR Prentice Hall, Upper Saddle River, N.J., 1999.

- [22] LJUNG, L., **Perspectives on System Identification**, Annual Reviews in Control vol. 34, April 2010.
- [23] VISCONTI, I. F., LIMA, D. A., COSTA, J. M. C. de S., SOBRINHO, N. R. de B. C., **Measurement-Based Load Modeling Using Transfer Functions for Dynamic Simulations**, IEEE Transactions on Power Systems, v. 29, p. 111-120, 2014.
- [24] CHIOU, C.-Y., *et al.*, **Development of a Micro-processor-based Transient Data Recording System for Load Behavior Analysis**, IEEE Transactions on Power Systems, vol. 8, No. 1, February, 1993.
- [25] MAITRA A., GAIKWAD A., POURBEIK P., BROOKS D., **Load Model Parameter Derivation Using an Automated Algorithm and Measured Data**, 2008 IEEE General Meeting.
- [26] BOX G., JENKINS, **Time Series Analysis: Forecasting & Control**, Prentice Hall Inc., 1994.
- [27] KARLSSON, D., HILL, D. J., **Modelling and identification of nonlinear dynamic loads in power systems**, *IEEE Trans. Power Syst.*, vol. 9, no. 1, pp. 157–166, Feb. 1994.
- [28] KOCK, J. A., *et al.* **Induction motor parameter estimation through an output error technique**, *IEEE Trans. Energy Convers.*, vol. 9, no. 1, pp. 69–76, Mar. 1994.
- [29] CHOI, B.-K. *et al.*, **Measurement-based dynamic load models: Derivation, comparison, and validation**, *IEEE Trans. Power Syst.*, vol. 21, no. 3, pp. 1276–1283, Aug. 2006.
- [30] JU, P., HANDSCHIN, E., KARLSSON, D., **Nonlinear dynamic load modelling: Model and parameter estimation**, *IEEE Trans. Power Syst.*, vol. 11, no. 4, pp. 1689–1697, Nov. 1996.
- [31] ROUHANI, A., ABUR, A., **Real-time dynamic parameter estimation for an exponential dynamic load model**, *IEEE Trans. Smart Grid*, vol. 7, no. 3, pp. 1530–1536, May 2016.
- [32] MIRANIAN A., ROUZBEHI, K., **Nonlinear power system load identification using local model networks**, *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2872–2881, Aug. 2013.

- [33] HIYAMA, T., TOKIEDA, M., HUBBI, W., ANDOU, H., **Artificial neural network based dynamic load modeling**, *IEEE Trans. Power Syst.*, vol. 12, no. 4, pp. 1576–1583, Nov. 1997.
- [34] WEN, J. Y., JIANG, L., WU, Q. H., CHENG, S. J., **Power system load modeling by learning based on system measurements**, *IEEE Trans. Power Del.*, vol. 18, no. 2, pp. 364–371, Apr. 2003.
- [35] LU, C. H., **Wavelet fuzzy neural networks for identification and predictive control of dynamic systems**, *IEEE Trans. Ind. Electron.*, vol. 58, no. 7, pp. 3046–3058, Jul. 2011.
- [36] KONTIS *et al.*: **Measurement-Based Dynamic Load Modeling Using The Vector Fitting Technique**, *IEEE Transactions On Power Systems*, Vol. 33, No. 1, January 2018.
- [37] MICHALEWICZ, Z., **Genetic algorithms + data structures = evolution programs**, Springer, 1992.
- [38] LI, Y. CHIANG, H., CHOI, B., CHEN, Y., HUANG, D., LAUBY, M., **"Load models for modeling dynamic behaviors of reactive loads: Evaluation and comparison"**, in *Electrical Power & Energy Systems* 30 (2008), 497-503, Elsevier Ltd.
- [39] KONTIS, E. O., PAPADOPOULOS T. A., CHRYSOCHOS, A. I., PAPAGIANNIS G. K., **Measurement-Based Dynamic Load Modeling**, *IEEE Transactions On Power Systems*, Vol. 33, No. 1, January 2018.
- [40] **Cepel, Programa Anatem – Análise de Transitórios Eletromecânicos – User Guide v10.04.03**, Brazil, 2010.
- [41] PARK, G. L., COLONY, R., **Voltage reduction as a means of reducing distribution load**, *IEEE Trans. Power App. Syst.*, vol. 96, pp.628 -634, 1977.
- [42] PREISS, R. F., WARNOCK, V. J., **Impact of volt-reduction on energy and demand**, *IEEE Trans. Power App. Syst.*, vol. 97, pp.1165 -1181, 1978.
- [43] EPRI, **Effects of Reduced Voltage on the Operation and Efficiency of Electric Loads**, Final Report EL-2036, September, 1981.
- [44] ERICKSON, J. C., GILLIGAN, S. R., **The Effects of Voltage reduction on Distribution Circuit Load**, *IEEE Power Engineering Review*, vol.2, Issue 7, July 1982.

- [45] WANG, Z., WANG, J., **Review on Implementation and Assessment of Conservation Voltage Reduction**, IEEE Transactions On Power Systems, Vol. 29, No. 3, May 2014.
- [46] AGUILÓ, M. D., *et al.*, **Field-Validated Load Model for the Analysis of CVR in Distribution Secondary Networks: Energy Conservation**, IEEE Transactions On Power Delivery, Vol. 28, No. 4, October 2013.
- [47] TANG X., HASAN K., MILANOVIĆ, J. V., Bailey K., Stott S., **Estimation and Validation of Characteristic Load Profile Through Smart Grid Trials in a Medium Voltage Distribution Network**, IEEE Transactions On Power Systems, Vol. 33, No. 2, March 2018.
- [48] LIU, H. J. *et al.*, **A Methodology to Analyze Conservation Voltage Reduction Performance Using Field Test Data**, 2014 IEEE International Conference on Smart Grid Communications.
- [49] SEN, P. K., LEE, K. H., **Conservation Voltage Reduction Technique: An Application Guideline for Smarter Grid**, IEEE Rural Electric Power Conference (REPC), Texas, USA, 2014.
- [50] EN 50160, **Voltage characteristics of electricity supplied by public distribution systems**, 1999.
- [51] National Institute of Standards and Technology - NIST, **Exploratory Data Analysis**, Available [Online]: <http://www.itl.nist.gov/div898/handbook//eda/eda.htm>.
- [52] TUKEY, JOHN, **Exploratory Data Analysis**, Addison-Wesley, 1977.
- [53] Electricity North West web site, **CLASS project** <http://www.enwl.co.uk/class>.
- [54] EPRI, **Green Circuit: Distribution Efficiency Case Studies**, EPRI, Technical Report, 2011. [Online] Available:
- [55] Electricity North West, **Trial Substation Selection Methodology**, [Online]. Available: August, 2013.
- [56] Electricity North West, **Design Approach to CLASS Trials and Associated test schedule**, [Online] Available.
- [57] SCHAFER, R. W., **What is a Savitzky-Golay filter**, IEEE Signal Processing Magazine, VOL. 28, No. 4, 2011.
- [58] ORFANIDIS, SOPHOCLES J., **Introduction to Signal Processing**. Englewood Cliffs, NJ: Prentice-Hall, 1996.

- [59] FISZ, MAREK, **Probability Theory and Mathematical Statistics**, John Wiley & Sons, 1963.
- [60] ELLIS, RICHARD B., **Statistical Inference – basic concepts**, Englewood Cliffs, NJ: Prentice-Hall, 1975.
- [61] GONICK, LARRY, **The cartoon guide to statistics**, HarperCollins Publishers, Inc, 1993.
- [62] BYRD, R. H., J. C. GILBERT, AND J. NOCEDAL. **A Trust Region Method Based on Interior Point Techniques for Nonlinear Programming**. Mathematical Programming, Vol 89, No. 1, 2000, pp. 149–185.
- [63] ARIF, A., WANG, Z., WANG J., MATHER B., BASHUALDO H., ZHAO D., **Load Modeling – A Review**, IEEE Transactions On Smart Grid, Vol. 9, No. 6, November 2018.