



UNIVERSIDADE ESTADUAL DE CAMPINAS  
Faculdade de Engenharia Elétrica e de Computação

Byron Alejandro Acuña Acurio

**Application of State Estimation and Machine  
Learning Techniques for the Energy Management of  
Microgrids**

**Aplicação de técnicas de estimação de estado e  
aprendizado de máquina para o gerenciamento  
energético das microrredes**

Byron Alejandro Acuña Acurio

**Application of State Estimation and Machine Learning Techniques for the Energy Management of Microgrids**

**Aplicação de técnicas de estimação de estado e aprendizado de máquina para o gerenciamento energético das microrredes**

Thesis presented to the School of Electrical and Computer Engineering of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Electrical Engineering, in the area of Electrical Energy.

*Tese apresentada à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor em Engenharia Elétrica, na área de Energia Elétrica.*

Supervisor/*Orientador*: Prof. Dr. Luiz Carlos Pereira da Silva

Co-supervisor/*Co-orientador*: Dr. Juan Camilo Lopez Amezquita

Este trabalho corresponde à versão final da tese defendida pelo aluno Byron Alejandro Acuña Acurio, e orientada pelo Prof. Dr. Luiz Carlos Pereira da Silva

---

Campinas  
2025

Ficha catalográfica  
Universidade Estadual de Campinas (UNICAMP)  
Biblioteca da Área de Engenharia e Arquitetura  
Rose Meire da Silva - CRB 8/5974

Ac93a Acuña Acurio, Byron Alejandro, 1986-  
Application of state estimation and machine learning techniques for the energy management of microgrids / Byron Alejandro Acuña Acurio. – Campinas, SP : [s.n.], 2025.

Orientador: Luiz Carlos Pereira da Silva.  
Coorientador: Juan Camilo Lopez Amezquita.  
Tese (doutorado) – Universidade Estadual de Campinas (UNICAMP), Faculdade de Engenharia Elétrica e de Computação.

1. Microrredes. 2. Aprendizado de máquina. 3. Sistemas de energia elétrica - Estimação de estados. 4. Tomada de decisões. I. Silva, Luiz Carlos Pereira da, 1972-. II. López Amézquita, Juan Camilo, 1989-. III. Universidade Estadual de Campinas (UNICAMP). Faculdade de Engenharia Elétrica e de Computação. IV. Título.

Informações complementares

**Título em outro idioma:** Aplicação de técnicas de estimação de estado e aprendizado de máquina para o gerenciamento energético das microrredes

**Palavras-chave em inglês:**

Microgrids

Machine learning

Electric power systems — State estimation

Decision making

**Área de concentração:** Energia Elétrica

**Titulação:** Doutor em Engenharia Elétrica

**Banca examinadora:**

Luiz Carlos Pereira da Silva [Orientador]

Carlos Alberto de Castro Júnior

Roberto Perillo Barbosa da Silva

Wendy Yadira Eras Herrera

Romis Ribeiro de Faissol Attux

**Data de defesa:** 14-03-2025

**Programa de Pós-Graduação:** Engenharia Elétrica

**Objetivos de Desenvolvimento Sustentável (ODS)**

ODS: 7. Energia acessível e limpa

**Identificação e informações acadêmicas do(a) aluno(a)**

- ORCID do autor: <https://orcid.org/0000-0002-6788-6833>

- Currículo Lattes do autor: <http://lattes.cnpq.br/1720607469550929>

Prof. Dr. Luiz Carlos Pereira da Silva (Presidente)

Prof. Dr. Carlos Alberto de Castro Júnior

Prof. Dr. Roberto Perillo Barbosa da Silva

Profa. Dra. Wendy Yadira Eras Herrera

Prof. Dr. Romis Ribeiro de Faissol Attux

A ata de defesa, com as respectivas assinaturas dos membros da Comissão Julgadora, encontra-se no SIGA (Sistema de Fluxo de Dissertação/Tese) e na Secretaria de Pós-Graduação da Faculdade de Engenharia Elétrica e de Computação.

# Acknowledgments

First, I extend my deepest gratitude to Professor Luiz Carlos Pereira da Silva, for the mentorship and support he has given me throughout my time at UNICAMP. I greatly admire his passion for both research and teaching, and it has been a true privilege to work under his supervision and guidance. He always offered a friendly and constructive environment for improving my work.

To my loving parents, Fanny and Edgar, and my brother Roberto, whose endless love and encouragement teach me day by day to be a better person. None of this would have been possible without their support.

To Diana, I am extremely fortunate to have you in my life and as my personal support. In the same way, thanks to all her family for always being aware of us.

I am grateful to my friends and colleagues who have accompanied me along this path.

This study was financed, in part, by the São Paulo Research Foundation (FAPESP), Brasil. Process Number #2020/03069-5, #2022/16881-5, and #2021/11380-5, Centro Paulista de Estudos da Transição Energética (CPTEn). This work is also developed under the Electricity Sector Research and Development Program PD-00063-3058/2019 - PA3058: “MERGE - Microgrids for Efficient, Reliable and Greener Energy”, regulated by the National Electricity Agency (ANEEL in Portuguese), in partnership with CPFL Energia (Local Electricity Distributor).

# Resumo

As microrredes são sistemas de energia sustentável integrados, capazes de operar tanto conectados à rede elétrica principal quanto de forma independente. Esses sistemas geralmente incorporam diversos recursos energéticos distribuídos (DERs), incluindo painéis fotovoltaicos, sistemas de armazenamento e geradores térmicos. No entanto, a natureza intermitente da geração de energia renovável em microrredes cria desafios estocásticos complexos para o controle em tempo real e a confiabilidade do sistema. Para enfrentar esses desafios, algoritmos avançados de aprendizado de máquina e técnicas de estimação de estado são essenciais para monitorar e responder a eventos dinâmicos e aleatórios fora do controle humano, permitindo um monitoramento preciso dos componentes da microrrede em diversos cenários operacionais.

A estimação de estado desempenha duas funções cruciais nas microrredes: converter leituras de medidores e dados disponíveis em estimativas confiáveis de variáveis não monitoradas e identificar erros significativos de medição. Por outro lado, o aprendizado de máquina pode auxiliar aos algoritmos de estimação de estado com previsões de variáveis não observáveis.

Este projeto de pesquisa foca no desenvolvimento e avaliação de ferramentas híbridas que combinam algoritmos de aprendizado de máquina e estimação de estado para aprimorar as funcionalidades do sistema de gestão de energia das microrredes baseado em dados.

**Palavras-chave:** Microrredes; Aprendizado de máquina; Estimação de estado; Abordagem orientada por dados.

# Abstract

Microgrids are integrated sustainable energy systems capable of operating both connected to and independently from the main power grid. These systems typically incorporate various distributed energy resources (DERs), including photovoltaic panels, storage systems, and thermal generators. However, the intermittent nature of renewable power generation in microgrids creates complex stochastic challenges for real-time control and system reliability. To address this, advanced machine learning algorithms and state estimation techniques are essential for monitoring and responding to dynamic, random events beyond human control, enabling accurate monitoring of microgrid components across various operational scenarios.

State estimation serves two crucial functions in microgrids: converting smart meter readings and available data into reliable estimates of unmonitored variables, and identifying significant measurement errors. Meanwhile, machine learning can aid state estimation algorithms with predictions of unobservable variables.

This research project focuses on developing and evaluating hybrid tools that combine machine learning and state estimation algorithms to enhance microgrid data-driven energy management functionalities.

**Key-words:** Microgrids; Machine learning; State estimation; Data-driven approach.

# List of Figures

2.1	Aerial view of the campus microgrid at the State University of Campinas (UNICAMP). . . . .	24
2.2	IEEE UKG-16 test microgrid. . . . .	25
2.3	Residual of estimated voltages of each phase using the WLS and WLAV models on Campusgrid-60 microgrid. . . . .	25
2.4	Residual of estimated voltages of each phase using the WLS and WLAV models on UKG-16 microgrid. . . . .	26
2.5	Sensitivity analysis of the WLS model for the Campusgrid-60 microgrid. . . . .	27
2.6	Sensitivity analysis of the WLS model for the UKG-16 microgrid. . . . .	27
3.1	Aerial view of the campus microgrid at the State University of Campinas (UNICAMP) [1]. . . . .	33
3.2	Proposed data event generator. . . . .	35
3.3	Proposed MLSE regularization process. . . . .	40
3.4	Recursive Average Model. . . . .	42
3.5	Workflow of the proposed MLSE. . . . .	43
3.6	L-curve for the proposed MLSE. $\lambda$ search space from $1e-10$ to $1e+10$ . . . . .	45
3.7	Test Model. . . . .	46
3.8	Results in terms of mean squared error of all Voltage magnitudes $V_{i,p}$ in grid-connected mode. . . . .	47
3.9	Results in terms of mean squared error of all Voltage phases $\theta_{i,p}$ in grid-connected mode. . . . .	47
3.10	Results in terms of mean squared error of all Voltage magnitudes $V_{i,p}$ in islanded mode. . . . .	48
3.11	Results in terms of mean squared error of all Voltage phases $\theta_{i,p}$ in islanded mode. . . . .	48
3.12	Test of the mean for all Voltage magnitudes $V_{i,p}$ . . . . .	49
3.13	Test of the mean for all Voltage phases $\theta_{i,p}$ . . . . .	49
3.14	Homogeneity test for all state variables. . . . .	50
4.1	(a) Original VGG-16 [2] . (b) Modified VGG-16 employed to perform a transfer learning task. . . . .	55
4.2	Example of an image from “transient attribute database $\mathcal{X}_t, \mathcal{Y}_t$ ”, which is publicly available in [3] with their attribute label. . . . .	56

4.3	A heat map shows that the modified VGG-16 neural network looks for very different points in the image, depending on the weather conditions. . . . .	57
4.4	Example of features extracted by convolutional layers. . . . .	57
4.5	Correlation Matrix of the 40 attributes obtained from the modified VGG-16 from the images. . . . .	58
4.6	Numerical results of the training modified VGG-16 neural network. . . . .	60
4.7	Comparison of the visual state estimated values and ground truth. . . . .	62
5.1	Proposed state of health estimation for lithium-ion batteries based on signal reconstruction. . . . .	67
5.2	Sensitivity analysis of signal reconstruction quality across various SNRs (10–50 dB). Our proposed approach based on Tikhonov consistently outperforms widely adopted signal processing methods, including the LASSO formulation, moving average (MA) filter, and the Kalman filter, across multiple accuracy metrics. Boxplots show (a) RMSE, (b) MAE, and (c) MAPE, with the proposed approach yielding the lowest error dispersion across all metrics, highlighting its effectiveness in preserving signal quality under noise. . . . .	78
5.3	Performance comparison of state-of-health (SoH) estimation across three scenarios using direct features reported in [4] vs. our proposed feature selection approach. In each scenario, the left plot shows predicted SoH at different numbers of aging cycles. The boxplots show RMSE, MAE, and MAPE. The proposed method consistently exhibits lower prediction errors, highlighting its effectiveness over direct features. . . . .	81
5.4	(a) In Scenario 1, our proposed approach consistently estimates the SoH throughout the battery life cycle. (b) In Scenario 2, the comparison highlights the robustness of our proposed approach against noisy data, whereas the DNN approach [5] is sensitive to noise. (c) In Scenario 3, the results demonstrate the resilience of our proposed approach to some missing values.	83
5.5	Comparison of proposed SoH approach (red) versus hybrid model (green) [6] across multiple scenarios based on NASA dataset [7] and detailed in Section 5.2.2. Left plots show the SoH predictions for each scenario along different numbers of aging cycles (the ground truth is in black). Boxplots of the corresponding error metrics RMSE, MAE, and MAPE. Overall, our proposed purely data-driven approach captures real-world battery aging dynamics more accurately and consistently than the hybrid model. . . . .	84
5.6	Comparison of proposed SoH approach (red) versus hybrid model (orange) [6], across multiple scenarios based on SNL dataset [8] and detailed in Section 5.2.2. Left plots show the SoH predictions for each scenario along different numbers of aging cycles (the ground truth is in black). Boxplots of the corresponding error metrics RMSE, MAE, and MAPE. We can see our proposed pure data-driven approach can adapt to the nuances of battery aging in ways the hybrid approach could not. . . . .	85

# List of Tables

2.1	Error of each method (for both systems) in percentage. . . . .	26
4.1	Comparison of the proposed modified VGG-16 neural network with previous works [3]. . . . .	61
5.1	Comparison of proposed SoH estimator with previous data-driven approaches.	66
5.2	Operating parameters of batteries from NASA dataset. . . . .	73
5.3	Operating parameters of batteries from SNL dataset. . . . .	74
5.4	Comparison of signal reconstruction accuracy for the proposed closed-form Tikhonov, the iterative LASSO formulation, a centred moving-average (MA) filter, and a recursive Kalman filter under five noise conditions (SNR = 10–50 dB). Lower values of RMSE, MAE, and MAPE indicate superior reconstruction quality. The best results in each column are in bold. . . . .	78
5.5	Sensitivity analysis of proposed features. Checkmark ( ) indicates inclusion and a dash (-) indicates exclusion of the corresponding proposed features: minimum discharge voltage ( $x_1$ ), time to minimum voltage ( $x_2$ ), minimum temperature at the start of discharge ( $x_3$ ), maximum discharge temperature ( $x_4$ ), and time elapsed between minimum and maximum temperature ( $x_5$ ). The best results are in bold. . . . .	80
5.6	Pearson correlation coefficients ( $r$ ) between each of the five proposed features and the battery state of health (SoH), reported separately for the training set, the independent test set, and the combined dataset. The strong positive correlations of the temporal features $x_2$ and $x_5$ with SoH corroborate their dominant explanatory power, whereas $x_1$ , $x_3$ , and $x_4$ show only weak or moderate association. A negative sign indicates an inverse monotonic relationship with SoH. . . . .	80
5.7	Comparison of proposed state-of-health (SoH) estimator with a deep neural network (DNN) using three handcrafted charging features, a Huber-regression model trained on ten direct statistics, and a hybrid estimator (physics-aided). Performance is evaluated under six aging scenarios, with results reported as root mean square error (RMS), mean absolute error (MAE), and mean absolute percentage error (MAPE). Lower values indicate higher accuracy. Aggregated results are labeled “Average”. The best results are in bold. . . . .	86

5.7	<i>Cont.</i> . . . . .	87
5.8	Comparative analysis of State of Health (SoH) estimation performance across six scenarios, demonstrating the effect of discharge C-rates, ambient temperatures, and charging protocols on RMSE, MAE, and MAPE metrics.	87

# List of Symbols

Symbol	Description
<i>Sets</i>	
$\Omega_b$	Set of nodes
$\Omega_l$	Set of branches
$\mathcal{F}$	Set of phases $\mathcal{F} = \{A, B, C\}$
$\Omega_b^V$	Set of voltage-magnitude measurements at the nodes
$\Omega_b^{QD}$	Set of reactive-power measurements at the nodes
<i>Indices</i>	
$i$	Node index, $i \in \Omega_b$
$ij$	Branch index, $ij \in \Omega_l$
$\phi, \psi$	Phase indices, $\phi, \psi \in \mathcal{F}$
$k$	Metered node index, $k \in \Omega_b$
<i>Parameters</i>	
$Z_{ij,\phi,\psi}$	Circuit impedance of branch $ij$ [ $\Omega$ ]
$R_{ij,\phi,\psi}$	Circuit resistance of branch $ij$ [ $\Omega$ ]
$X_{ij,\phi,\psi}$	Circuit reactance of branch $ij$ [ $\Omega$ ]
$Z'_{ij,\phi,\psi}$	Transformed impedance: $Z'_{ij,\phi,\psi} = Z_{ij,\phi,\psi} (\theta_\psi - \theta_\phi)$
$V_{\min}$	Minimum voltage magnitude [kV]
$V_{\max}$	Maximum voltage magnitude [kV]
$W_{k,\phi}^V$	Weight of voltage-magnitude measurement at node $k$ , phase $\phi$
$W_{k,\phi}^{PD}$	Weight of active-power measurement at node $k$ , phase $\phi$
$W_{k,\phi}^{QD}$	Weight of reactive-power measurement at node $k$ , phase $\phi$
$V_{k,\phi}^{\text{med}}$	Measured voltage magnitude at node $k$ , phase $\phi$
$P_{k,\phi}^{D,\text{med}}$	Measured active-power demand at node $k$ , phase $\phi$
$Q_{k,\phi}^{D,\text{med}}$	Measured reactive-power demand at node $k$ , phase $\phi$
<i>Variables</i>	
$e_{k,\phi}^V$	Residual of voltage-magnitude measurement at node $k$ , phase $\phi$
$e_{k,\phi}^{PD}$	Residual of active-power demand measurement at node $k$ , phase $\phi$
$e_{k,\phi}^{QD}$	Residual of reactive-power demand measurement at node $k$ , phase $\phi$
$V_{i,\phi}$	Voltage magnitude at node $i$ , phase $\phi$ [kV]
$P_{i,\phi}^D$	Active-power demand at node $i$ , phase $\phi$ [kW]
$Q_{i,\phi}^D$	Reactive-power demand at node $i$ , phase $\phi$ [kVAr]
$P_{ij,\phi}^L$	Active-power loss in branch $ij$ , phase $\phi$ [kW]
$Q_{ij,\phi}^L$	Reactive-power loss in branch $ij$ , phase $\phi$ [kVAr]
$P_{i,\phi}^S$	Active-power injection at node $i$ , phase $\phi$ [kW]
$Q_{i,\phi}^S$	Reactive-power injection at node $i$ , phase $\phi$ [kVAr]

# Contents

<b>1</b>	<b>Introduction</b>	<b>15</b>
1.1	Motivation . . . . .	15
1.2	Challenges . . . . .	15
1.3	Aims of this thesis . . . . .	16
1.4	Summary of Contributions and Thesis Outline . . . . .	17
<b>2</b>	<b>Contribution I</b>	<b>19</b>
2.1	Introduction . . . . .	20
2.2	Optimal unbalanced three-phase state estimation based on WLS criteria .	21
2.3	Optimal unbalanced three-phase state estimation based on WLAV criteria	22
2.4	Test System 1 (Campusgrid-60) . . . . .	23
2.5	Test System 2 (UKG-16) . . . . .	23
2.6	Tests and Results . . . . .	23
	2.6.1 Quality . . . . .	24
2.7	Computational Efficiency . . . . .	25
2.8	Conclusions . . . . .	27
<b>3</b>	<b>Contribution II</b>	<b>29</b>
3.1	Introduction . . . . .	30
3.2	Classical State Estimation . . . . .	32
3.3	Data Generation . . . . .	34
	3.3.1 Heteroscedastic noise measurements . . . . .	35
3.4	Proposed MLSE and Regularization Process . . . . .	36
	3.4.1 Machine Learning Model . . . . .	37
	3.4.2 Least-Square Optimization Model . . . . .	37
	3.4.2.1 Application of the Tikhonov Regularization . . . . .	39
	3.4.3 Recursive Average Model . . . . .	42
3.5	Materials and Methods . . . . .	42
	3.5.1 Statistical analysis . . . . .	43
3.6	Numerical Results . . . . .	44
	3.6.1 Computational complexity analysis . . . . .	46
3.7	Conclusions and Discussion . . . . .	51

<b>4</b>	<b>Contribution III</b>	<b>52</b>
4.1	Introduction . . . . .	53
4.2	Proposed Approach . . . . .	54
4.2.1	Transfer Learning Procedure . . . . .	54
4.2.2	Intermediate Representation . . . . .	57
4.2.3	Binary Hypothesis Test . . . . .	58
4.3	Experimental Setup . . . . .	59
4.4	Results And Discussion . . . . .	61
4.5	Conclusions . . . . .	62
4.6	Future Works . . . . .	62
<b>5</b>	<b>Contribution IV</b>	<b>63</b>
5.1	Introduction . . . . .	64
5.2	Materials and Methods . . . . .	67
5.2.1	Proposed Approach . . . . .	67
5.2.1.1	Proposed Signal Reconstruction stage . . . . .	68
5.2.1.2	Proposed Feature Extraction stage . . . . .	70
5.2.1.3	State of Health (SoH) Estimation . . . . .	71
5.2.2	Datasets . . . . .	72
5.2.2.1	NASA Data Overview . . . . .	72
5.2.2.2	SNL Data Overview . . . . .	73
5.2.3	Accuracy Evaluation Metrics . . . . .	74
5.3	Numerical Results and Discussions . . . . .	75
5.3.1	Experimental Setup . . . . .	75
5.3.2	Sensitivity Analysis of Signal Reconstruction . . . . .	76
5.3.3	Sensitivity and Correlation Analysis of Proposed Features . . . . .	79
5.3.4	Comparison of Feature Selection: Direct Features vs Proposed . . . . .	81
5.3.5	Data-Driven Robustness Comparison . . . . .	81
5.3.5.1	DNN . . . . .	82
5.3.6	Comparison with Hybrid Model . . . . .	83
5.4	Conclusions . . . . .	88
<b>6</b>	<b>Conclusion and Future Work</b>	<b>89</b>
6.1	Conclusions . . . . .	89
6.2	Future Works . . . . .	90
	<b>References</b>	<b>91</b>
	<b>Appendix A Permission Grants</b>	<b>106</b>
	<b>Appendix B List of Publications</b>	<b>107</b>

# Chapter 1

## Introduction

### 1.1 Motivation

Electric power systems have historically been characterized by a one-way electricity flow from centralized generation sources transmitted over long distances to meet consumer demand [9]. However, the global transition towards clean and affordable electrical energy is a pressing necessity in the face of climate change and the depletion of fossil fuel resources [10].

Microgrids have emerged as a promising solution to minimize losses, integrate renewable energy sources (RES), and create innovative business opportunities through flexible distributed energy resources. The global microgrid market is expected to reach \$30.9 billion by 2027 [11]. Thus, the field of microgrid research holds significant potential for growth and innovation.

Microgrid monitoring systems are essential for real-time control of RES [12]. To do this, state estimation and machine learning methods determine the most likely electrical state across various operating conditions [13], even with limited available measurements.

This becomes crucial given the inherent uncertainty of RES, which introduces additional complexity to electricity flows and operations. The challenge intensifies with data-related issues, including redundancy, duplication, errors, missing data [14], and high dimensionality.

Hence, there is a clear need to develop integrated methods using machine learning and state estimation algorithms to address the aforementioned challenges, using a new paradigm named data-driven in the microgrid field[15].

### 1.2 Challenges

Traditional state estimation methods developed for transmission systems cannot be directly applied to microgrids, due to fundamental differences in their characteristics [16].

For instance, microgrids experience significant voltage fluctuations and bidirectional power flows [17] Furthermore, microgrids have the following challenges:

**System Imbalance:** Unlike transmission networks, microgrid systems operate under unbalanced conditions, necessitating three-phase modeling for accurate analysis [18].

**Branch Parameters** In transmission systems, high X/R ratios enable decoupled estimation of voltage magnitudes and angles. In contrast, microgrids with low X/R ratios cannot use these simplifications and require more precise phase angle measurements [19].

**Model Uncertainty:** Microgrids can use various energy sources, such as solar panels, small wind turbines, batteries, and conventional generators. The choice of energy sources often depends on the geographical location, the availability of resources, and specific energy needs [20]. However, solar energy is generally adopted. Hence, weather-dependent generation patterns (uncertainties) can dramatically impact the amount of peak power supplied on any given day.

**Observability Challenges:** Microgrid systems are inherently less observable than transmission networks, with fewer measurement points relative to system size [21].

**Smart Meters:** Unlike transmission networks, where phasor measurement units (PMU) are ubiquitous, microgrids generally make use of low-cost power meters, with sampling rates ranging from minutes to hours to monitor key electrical variables such as active and reactive power [22].

## 1.3 Aims of this thesis

Based on the microgrid challenges described in Section 1.2, this thesis aims to study and implement machine learning and state estimation techniques dedicated to microgrids. These methods can be used as data processing tools and estimators for a data-driven energy management system. These techniques should be suitable for both grid-connected and islanded operations, providing useful information for real-time microgrid operation.

The specific objectives are:

Research the need for static state estimators for microgrids.

Research and develop state estimation methodologies for microgrids in both grid-connected and islanded modes.

Develop a machine learning method to reduce renewable generation uncertainty error in microgrids (false data in renewable generation)

Develop a machine learning method to reduce uncertainty error in battery state of health employed in microgrids.

## 1.4 Summary of Contributions and Thesis Outline

This thesis is presented as a compilation of four papers, which are appended in Chapters 2 through 5. The contents presented here are a faithful representation of the published materials. In Chapter 6, we conclude this work and discuss various of our proposed data-driven applications for energy management systems of microgrids and encourage future research.

**Chapter 2** contains the first contribution of this work, entitled “State Estimation for Unbalanced Three-Phase AC Microgrids Based on Mathematical Programming,” which was published in IEEE ISGT 2023. This work presents a novel approach for state estimation in unbalanced three-phase AC microgrids using mathematical programming. The first key contribution is the development of a new state estimation model based on mathematical programming that can effectively incorporate available measurements as constraints. This model is suitable for three-phase unbalanced and asymmetric distribution microgrids with general topologies, addressing cases that would be unobservable using traditional approaches based on Newton-like algorithms. The proposed methodology is notably flexible and can be adapted to different unbalanced three-phase AC microgrids. Additionally, the work includes a comprehensive sensitivity study comparing state estimation formulations using both WLS and WLAV criteria. Through testing on both a real microgrid at the State University of Campinas and an IEEE test system, the results demonstrate that while both methods achieve high accuracy, the WLS-based estimator offers superior computational efficiency compared to WLAV.

**Chapter 3** contains the second contribution of this work, entitled “Design and Implementation of a Machine Learning State Estimation Model for Unobservable Microgrids,” which was published in IEEE Access 2022. This work introduces a novel machine learning state estimation (MLSE) model designed for unobservable and unbalanced three-phase AC microgrids operating under heteroscedastic noise conditions. The approach innovates by eliminating the need for pseudo-measurements and prior knowledge of measurement variances. Instead, a compact training dataset from OpenDSS power flow simulations was employed to fit a data-driven model. At its core, the method employs a new Tikhonov regularization operator to capture state variable-measurement relationships, complemented by a recursive average model that effectively handles varying measurement noise. Validated using real data from the University of Campinas (UNICAMP) microgrid, the model achieved high accuracy with mean squared errors of approximately  $1e-5$  outperforming traditional weighted least squares estimation  $1e-3$  benchmark. The approach can be used in diverse microgrid configurations while maintaining robust performance in both grid-connected and islanded operations.

**Chapter 4** presents the third contribution of this work, entitled “Visual State Estimation for False Data Injection Detection of Solar Power Generation,” which was published in the journal *Engineering Proceedings*. This work introduces a novel approach to detecting false data injection attacks (FDIAs) in solar power generation systems using visual data and machine learning. The method employs a modified VGG-16 neural network to analyze sky images and obtain intermediate representations that provide both textual and numerical explanations about visual weather conditions. The approach innovates by creating a mapping between visual atmospheric conditions and expected power generation, allowing detection of anomalies that could indicate FDIAs. The modified VGG-16 architecture achieved superior accuracy compared to previous methods, with a mean squared error of 0.0319 versus 0.043 for the next best approach. The proposed approach can identify tampered measurements through a proposed binary hypothesis test framework using a chi-square distribution, providing power system operators with an additional layer of security against cyber attacks.

**Chapter 5** presents the fourth contribution of this work, entitled “Robust Data-Driven State of Health Estimation of Lithium-Ion Batteries Based on Reconstructed Signals,” which was published in the journal *Energies*. This chapter demonstrates that robustness in state of health (SoH) estimation for lithium-ion batteries is influenced by the choice of machine learning architecture, loss function, feature selection, and signal reconstruction technique. In addition, we found that tracking the time to minimum discharge voltage and the time to maximum discharge temperature can be used as effective features to estimate SoH in data-driven models, as they are directly correlated with capacity loss and a decrease in power output. Finally, we introduce a novel approach SoH estimation method of Lithium-Ion Batteries Based on Reconstructed Signals. Our proposed approach outperforms previous state-of-the-art machine learning models in terms of accuracy in noisy conditions, using small training datasets.

**Chapter 6** presents the conclusions of this thesis and recommends possible directions for future research.

# Chapter 2

## Contribution I

### State Estimation for Unbalanced Three-Phase AC Microgrids Based on Mathematical Programming<sup>1</sup>

**Abstract:** State estimation (SE) helps to determine the most-likely steady-state operation of microgrids based on field measurements. It is a fundamental data processing tool responsible for supporting and increasing system visibility and filtering errors that may appear in real-time measurements, system topology, and parameters. Hence, it serves as a reliable basis for the energy management system (EMS) infrastructure. Existing SE methods are based on Newton-like algorithms rather than using specific optimization. Nevertheless, SE can always be reduced to a constrained optimization problem, with the objective of minimizing a given criterion, e.g., weighted least squares (WLS) or weighted least absolute value (WLAV). In this context, this paper presents two unbalanced three-phase AC SE methods based on mathematical programming for microgrids. The effectiveness and validity of the proposed state estimation approach are demonstrated on a real microgrid located at the State University of Campinas, in Brazil, and over one widely known IEEE test system. The method can be easily adapted to other microgrids with different configurations, distributed energy resources, and measurements. Results show that the proposed methods report high accuracy, but the state estimator based on WLS is faster than the WLAV.

**Keywords:** Microgrids, state estimation, nonlinear programming problem, unbalanced three-phase AC network.

---

<sup>1</sup>Contribution I is published as: **Byron Alejandro Acuña Acurio**, Diana Estefanía Chérrez Baragán, Juan Camilo López, Felipe Grijalva, Juan Carlos Rodríguez and Luiz Carlos Pereira da Silva, "State Estimation for Unbalanced Three-Phase AC Microgrids Based on Mathematical Programming" [23] in IEEE ISGT 2023, doi: <https://doi.org/10.1109/ISGT51731.2023.10066353>

## 2.1 Introduction

State estimation is one of the most important data processing tools for information retrieval in power systems [24]. When properly implemented, it identifies the value of the state variables of an electrical grid from available measurements in real-time. Therefore, state estimation plays an important role in modern energy management systems (EMS) providing a complete, accurate, consistent, and reliable database [25]. Classic state estimation methods for monitoring microgrids are centralized [26] using asynchronous measurements [27] because consumption and generation generally are transmitted in time intervals of around 15 to 60 mins. using low-cost smart-meters instead of phase measurement units (PMUs).

Currently, the WLS method based on Newton-like algorithms is the most frequently used for SE in power systems [28]. This method minimizes the square of the sum of weighted residuals, and its main problem is that the gain matrix may be ill-conditioned, which is often the case for state estimation in unbalanced three-Phase AC microgrids. If this happens, the solution may not converge and the system states cannot be accurately obtained [29], [30]. The numerically ill-conditioned problem has been solved by the confidence region method with quadratic regulation factoring (QR), but the convergence problem still exists [31].

Given the fundamental role of SE in a microgrid, two SEs using mathematical programming are proposed in this work. One is based on weighted least squares (WLS) and the other on weighted least absolute value (WLAV). Their main advantages are that (1) they do not require measurement conversion, and (2) they can be easily implemented using commercial solvers to admit any combination of measurements as constraints, where the explicit optimization variables are the measurement residuals. The developed models are suitable for three-phase unbalanced and asymmetric distribution microgrids of general topologies. This latter case of study is unobservable<sup>2</sup> without using pseudo-measurements for traditional WLS and WLAV approaches based on Newton-like algorithms. A solution to this problem can be obtained using the models proposed in this study. The main contributions are as follows:

A new unbalanced three-phase AC microgrid state estimation model based on mathematical programming, in which available measurements can be incorporated as restrictions.

The state estimation model is formulated with weighted least squares (WLS) and weighted least absolute value (WLAV) to perform a sensitivity study.

---

<sup>2</sup>A microgrid is considered unobservable when there are more state variables than available measurements.

Finally, it is worth mentioning that the proposed methodology is flexible and can be applied to different unbalanced three-phase AC microgrids.

## 2.2 Optimal unbalanced three-phase state estimation based on WLS criteria

In this section, the mathematical model to solve the unbalanced three-phase state estimation based on WLS criteria is presented. The proposed model below was implemented with AMPL [32] using the IPOPT [33] solver and was tested with Campusgrid-60 and UKG-16 IEEE test networks respectively.

The objective function shown in equation (2.1) minimizes the sum of weighted squared residuals [34]. In this case, state estimation was implemented with the following measurement types: (i) voltage magnitude; and (ii) active and (iii) reactive power demands.

$$\min \left\{ \sum_{k \in \Omega_b^V} \sum_{\phi \in \mathcal{F}} e_{k,\phi}^V W_{k,\phi}^V e_{k,\phi}^V + \sum_{k \in \Omega_b^{PD}} \sum_{\phi \in \mathcal{F}} e_{k,\phi}^{PD} W_{k,\phi}^{PD} e_{k,\phi}^{PD} + \sum_{k \in \Omega_b^{QD}} \sum_{\phi \in \mathcal{F}} e_{k,\phi}^{QD} W_{k,\phi}^{QD} e_{k,\phi}^{QD} \right\} \quad (2.1)$$

subject to:

$$P_{ij,\phi}^L = \sum_{\psi \in \mathcal{F}} \frac{1}{V_{i,\phi} V_{i,\psi}} \left[ R'_{ij,\phi,\psi} (P_{ij,\phi} P_{ij,\psi} + Q_{ij,\phi} Q_{ij,\psi}) + X'_{ij,\phi,\psi} (Q_{ij,\psi} P_{ij,\phi} - P_{ij,\psi} Q_{ij,\phi}) \right]; \quad \forall ij, \phi \quad (2.2)$$

$$Q_{ij,\phi}^L = \sum_{\psi \in \mathcal{F}} \frac{1}{V_{i,\phi} V_{i,\psi}} \left[ R'_{ij,\phi,\psi} (P_{ij,\psi} Q_{ij,\phi} - Q_{ij,\psi} P_{ij,\phi}) + X'_{ij,\phi,\psi} (P_{ij,\psi} P_{ij,\phi} + Q_{ij,\psi} Q_{ij,\phi}) \right], \quad \forall ij, \phi \quad (2.3)$$

$$\sum_{ki \in \Omega_l} P_{ki,\phi} - \sum_{ij \in \Omega_l} (P_{ij,\phi} + P_{ij,\phi}^L) + \sum_{i \in \Omega_b} P_{i,\phi}^S = P_{i,\phi}^D \quad \forall i \in \Omega_b, \forall \phi \in \mathcal{F} \quad (2.4)$$

$$\sum_{ki \in \Omega_l} Q_{ki,\phi} - \sum_{ij \in \Omega_l} (Q_{ij,\phi} + Q_{ij,\phi}^L) + \sum_{i \in \Omega_b} Q_{i,\phi}^S = Q_{i,\phi}^D \quad \forall i \in \Omega_b, \forall \phi \in \mathcal{F} \quad (2.5)$$

$$\begin{aligned}
V_{i,\phi}^2 - V_{j,\phi}^2 &= 2 \sum_{\psi \in \mathcal{F}} \left( R'_{ij,\phi,\psi} P_{ij,\psi} + X'_{ij,\phi,\psi} Q_{ij,\psi} \right) \\
&- \frac{1}{V_{i,\phi}^2} \left[ \left( R'_{ij,\phi,\psi} P_{ij,\psi} + X'_{ij,\phi,\psi} Q_{ij,\psi} \right) \right]^2 \\
&- \frac{1}{V_{i,\phi}^2} \left[ \left( R'_{ij,\phi,\psi} Q_{ij,\psi} + X'_{ij,\phi,\psi} P_{ij,\psi} \right) \right]^2, \forall ij \in \Omega_l, \forall \phi \in \mathcal{F}
\end{aligned} \tag{2.6}$$

$$P_{k,\phi}^{D,med} = P_{k,\phi}^D + e_{k,\phi}^{PD} \quad \forall k \in \Omega_b^{PD}, \forall \phi \in \mathcal{F} \tag{2.7}$$

$$Q_{k,\phi}^{D,med} = Q_{k,\phi}^D + e_{k,\phi}^{QD} \quad \forall k \in \Omega_b^{QD}, \forall \phi \in \mathcal{F} \tag{2.8}$$

$$V_{k,\phi}^{med} = V_{k,\phi} + e_{k,\phi}^V \quad \forall k \in \Omega_b^V, \forall \phi \in \mathcal{F} \tag{2.9}$$

$$V_{min} \leq V_{i,\phi} \leq V_{max} \quad \forall i \in \Omega_b, \forall \phi \in \mathcal{F} \tag{2.10}$$

$$0 \leq P_{i,\phi}^D \quad \forall i \in \Omega_b, \forall \phi \in \mathcal{F} \tag{2.11}$$

$$0 \leq Q_{i,\phi}^D \quad \forall i \in \Omega_b, \forall \phi \in \mathcal{F} \tag{2.12}$$

$$0 \leq P_{i,\phi}^S \quad \forall i \in \Omega_b, \forall \phi \in \mathcal{F} \tag{2.13}$$

$$0 \leq Q_{i,\phi}^S \quad \forall i \in \Omega_b, \forall \phi \in \mathcal{F} \tag{2.14}$$

Constraints (2.2) and (2.3) calculate the active and reactive power losses in each circuit, as it was presented in [35], [36]. On the other hand (2.4) and (2.5) are constraints for the balance of active and reactive power in the microgrid. In order to model voltage magnitude drop in lines, equation (2.6) was employed, as presented in [35], [36]. Constraints (2.7), (2.8) and (2.9) calculate the residuals of the measurements of active and reactive power demands, and voltage magnitude, respectively. The voltage limits are guaranteed by (2.10). Finally, the domain restrictions of the decision variables (so that the solution is always based in positive values of power) are given in (2.11), (2.12), (2.13), and (2.14).

## 2.3 Optimal unbalanced three-phase state estimation based on WLAV criteria

The aforementioned SE based on WLS criteria can be easily adapted to WLAV just by replacing the objective function of equation (2.1) with equation (2.15). State estimation based on WLAV criteria was implemented with AMPL [32] using the solvers SNOPT [37] and Knitro [38] to resolve the test cases of the Campusgrid-60 and UKG-16 microgrids, respectively.

Objective function (2.15) minimizes the sum of the absolute value of weighted residuals [39]. In this case, SE was implemented with the same measurement types as in the

WLS criteria: (i) voltage magnitude; and (ii) active and (iii) reactive power demands.

$$\min \left\{ \sum_{k \in \Omega_b^V} \sum_{\phi \in \mathcal{F}} W_{k,\phi}^V |e_{k,\phi}^V| + \sum_{k \in \Omega_b^{PD}} \sum_{\phi \in \mathcal{F}} W_{k,\phi}^{PD} |e_{k,\phi}^{PD}| \right. \\ \left. + \sum_{k \in \Omega_b^{QD}} \sum_{\phi \in \mathcal{F}} W_{k,\phi}^{QD} |e_{k,\phi}^{QD}| \right\} \quad (2.15)$$

## 2.4 Test System 1 (Campusgrid-60)

The University of Campinas (UNICAMP) is currently deploying a project [1] that transforms a part of its main campus into a modern microgrid named Campusgrid-60. Fig. 2.1 shows an aerial view of the microgrid and the main components. The optimal unbalanced three-phase state estimators based on WLS and WLAV criteria were tested in Campusgrid-60. It must be considered that although the tested system has 60 nodes, many of them are medium-voltage stepping or transfer nodes that do not have net power injection. On the other hand, the load nodes have transformers with meters on their secondary sides. Campusgrid-60 is an interesting case study because it is intended to test the ability of the model to obtain the state of all 60 three-phase nodes using only a single voltage measurement at the common coupling point (PCC), and 66 measurements of active and reactive power from smart meters on the secondary of the transformers. The nominal voltage of the Campusgrid-60 microgrid is  $V_{nom} = 11.9$  kV in the primary network, and supplies various buildings on the university campus through distribution Delta/Wye-G connected transformers with nominal capacities varying between 15 kVA and 1000 kVA. Smart meters at the low voltage side of these transformers measure the three-phase real and reactive powers, currents, and voltages in real time with a fifteen-minute sampling rate, enabling data analysis and state estimation.

## 2.5 Test System 2 (UKG-16)

A simplified single-line diagram of an unbalanced three-phase UKG-16 microgrid is shown in Fig. 2.2. This case of study is interesting because the unbalance causes droop voltages in some nodes.

## 2.6 Tests and Results

The formulation for the unbalanced three-phase SE problem presented in Sections 2.2 and 2.3 was validated on CampusGrid-60 microgrid at the State University of Campinas (UNICAMP) and on an IEEE UKG-16 test network. For both systems, only one condition

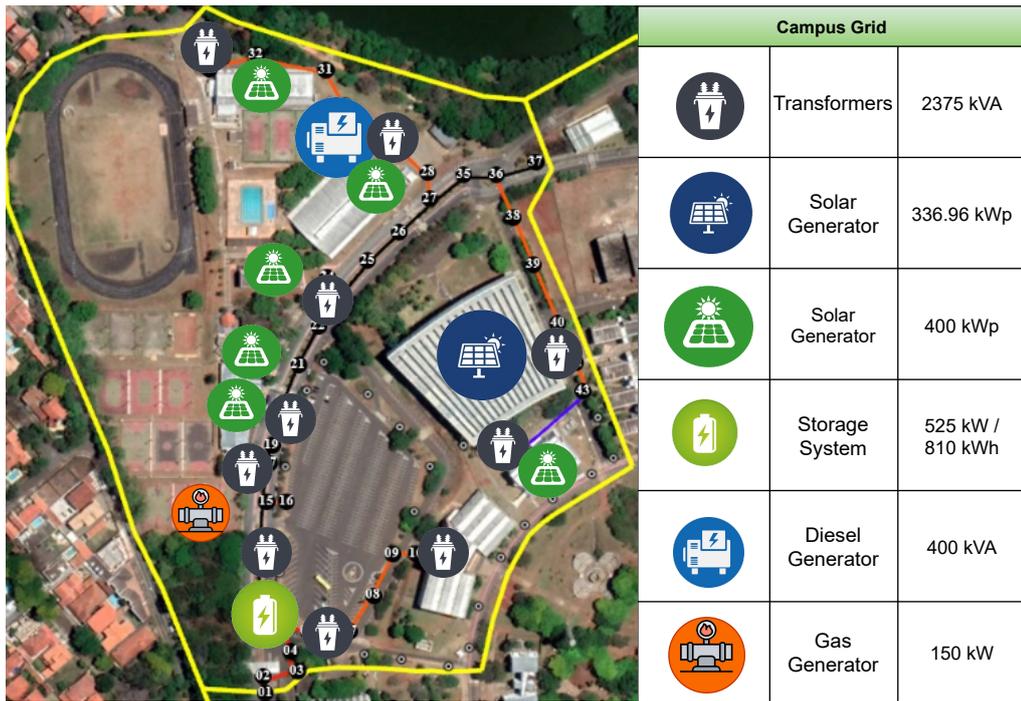


Figure 2.1: Aerial view of the campus microgrid at the State University of Campinas (UNI-CAMP).

of load, generation, and topology was considered. The non-linear programming (NLP) models were implemented in the mathematical programming language AMPL [32], and solved through commercial solvers. In the case of WLS, the solver IPOPT [33] was used on the two test microgrids. In the case of WLAV, the IPOPT solver did not converge. Thus, the solvers SNOPT [37] and Knitro [38] were used instead on the CampusGrid and UKG-16 microgrids, respectively. All tests were performed using a workstation with an Intel(R) Core(TM) i7 870 processor and 8 GB RAM.

### 2.6.1 Quality

Fig. 2.3 shows residuals of estimated voltages for each case considered. It can be seen that for the Campusgrid-60 test microgrid, both the WLS and WLAV models present a similar performance in terms of precision. Both state estimators report high performance because residuals are near zero while the estimations are in kilo-volts. However, the WLAV model is computationally much more cumbersome than the WLS model, as is shown in Fig. 2.5.

Fig. 2.4 shows residuals of estimated voltages for each case considered for the UKG-16 test network. In this test, it was necessary to change the initial point from 1 to 1.04 per unit to converge. In addition, it was necessary to swap solvers from SNOPT to Knitro, as the IPOPT[33] did not converge. In terms of precision, it can be seen in Fig. 2.6 that WLAV has the same performance as WLS but with a higher computational cost. Table 2.1 shows the error of each method (for both systems) in percentage.

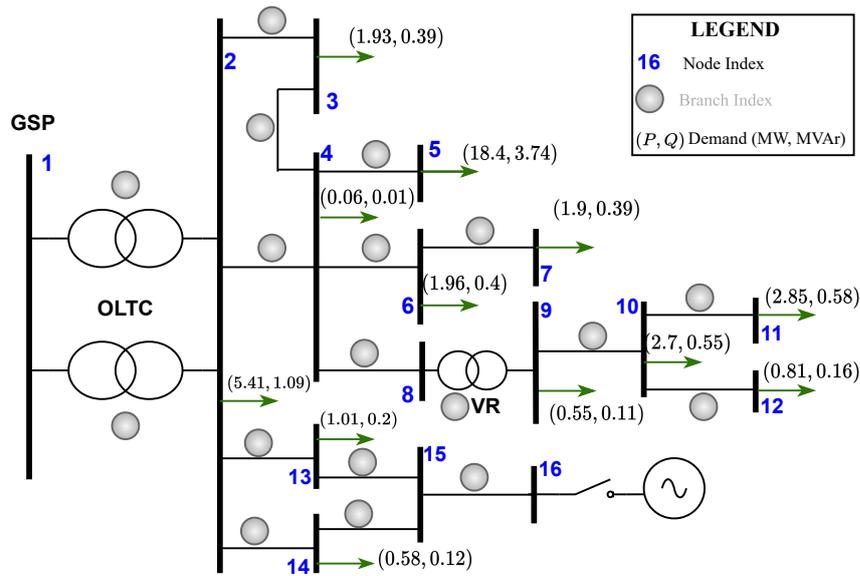


Figure 2.2: IEEE UKG-16 test microgrid.

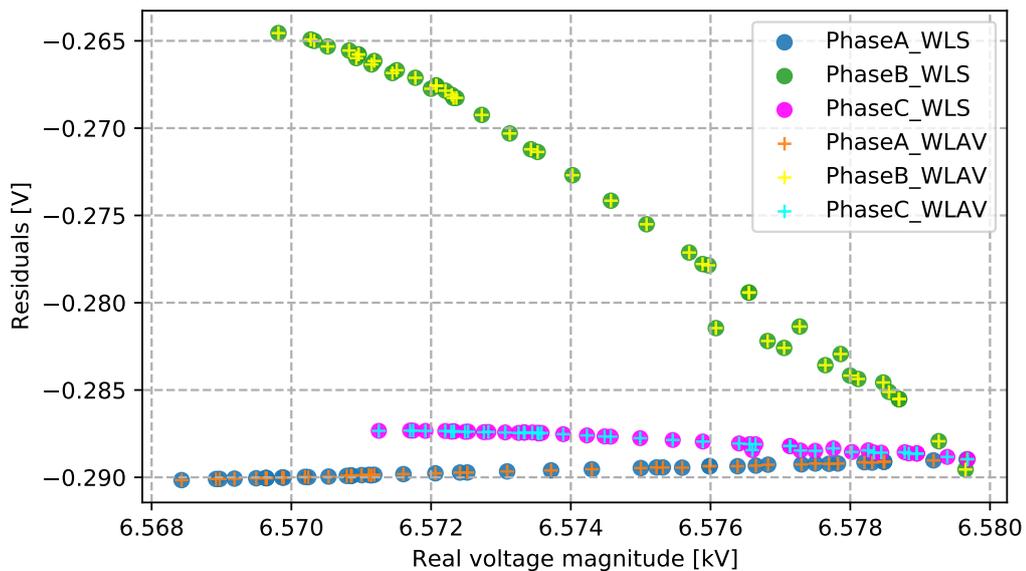


Figure 2.3: Residual of estimated voltages of each phase using the WLS and WLAV models on Campusgrid-60 microgrid.

## 2.7 Computational Efficiency

In this section, the convergence process of the state estimators presented above is tested to observe whether it improves or worsens the required number of iterations when new voltage measurements are added.

Fig. 2.5 illustrates, for each case considered, the number of iterations required when voltage measurements are incorporated in state estimation on Campusgrid-60 test case. Initially, only the voltage measurements of the Point of Common Coupling (PCC) were used. These are three corresponding measurements (one for each phase). Then the measurements from the smart meters on the transformer's secondary were incorporated

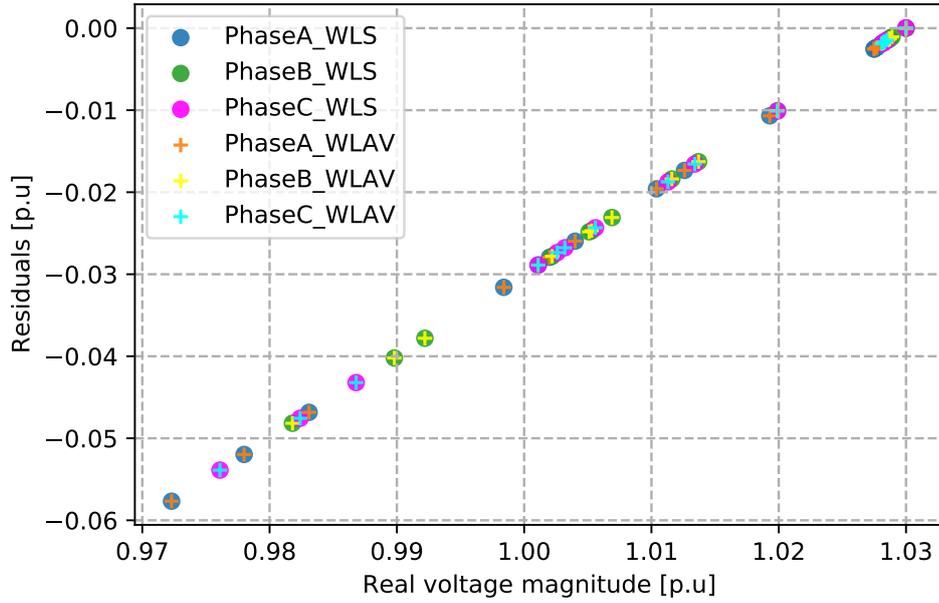


Figure 2.4: Residual of estimated voltages of each phase using the WLS and WLAV models on UKG-16 microgrid.

Table 2.1: Error of each method (for both systems) in percentage.

Microgrids	Measurements	Error(%)		
		Max	Average	Min
Campusgrid-60	$V_k, \text{WLS}$	0.0044	0.0042	0.00405
	$V_k, \text{WLAV}$	0.0044	0.0042	0.00405
UKG-16	$V_k, \text{WLS}$	0.055	0.0275	0.0001
	$V_k, \text{WLAV}$	0.055	0.0275	0.0001

to observe the behavior of the models. It can be seen that the best performance is obtained using only the PCC measurements (of each phase), with four iterations to solve the non-linear problem. On the other hand, when considering all the low-voltage measurements from the smart meters, the model takes longer to converge (approximately 9 iterations) in the WLS case. On the other hand, WLAV needs a minimum of 1786 iterations using only the PCC measurements and 2714 iterations in the worst-case scenario.

Fig. 2.6 presents the results of increasing the number of voltage measurements in the WLS and WLAV models, on the UKG-16 test case. Again, only the PCC voltage measurements were considered initially. Measurements on bus 4 (transfer or step) and on bus 15 were also considered. The sensitivity results of the WLS model in the UKG-16 test network presented a different behavior from what was observed in the Campusgrid-60 network. In this case, the WLS model showed greater sensitivity, as the required iterations increased from 3 to 45 i.e. a variation of 42 iterations. This is much larger than what was observed in the Campusgrid-60 test case, where the worst-case iteration change was 6. While WLAV in this test case has less sensitivity to the increase in measurements, it

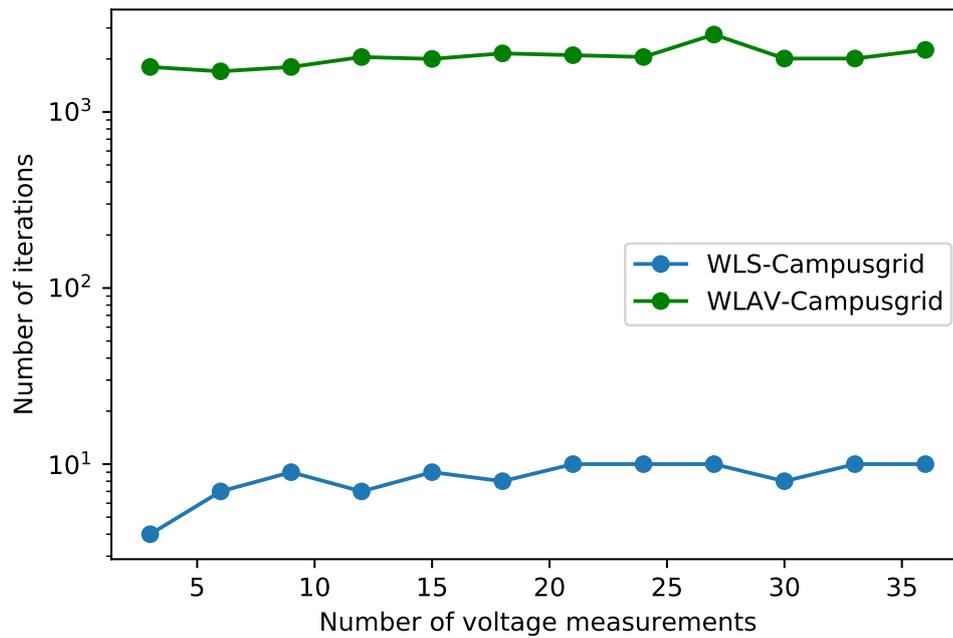


Figure 2.5: Sensitivity analysis of the WLS model for the Campusgrid-60 microgrid.

can be seen that in turn, it needs a minimum of 218 and a maximum of 266 iterations to converge. This is much more than the worst case for the WLS (45 iterations).

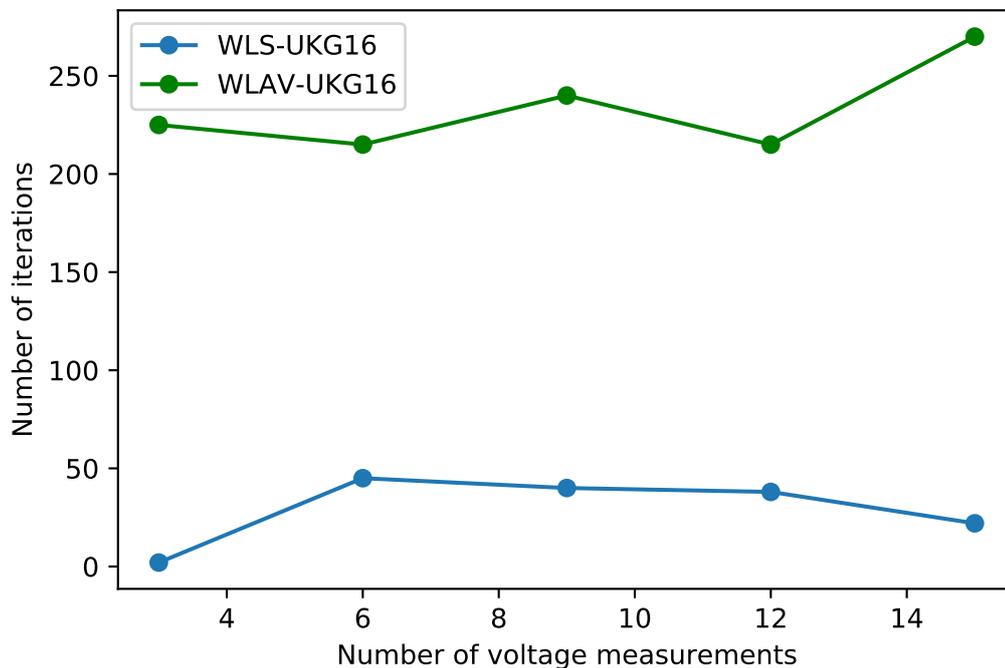


Figure 2.6: Sensitivity analysis of the WLS model for the UKG-16 microgrid.

## 2.8 Conclusions

In this work two SEs, one based on weighted least squares (WLS) and another one based on weighted least absolute value (WLAV) criteria, were modeled using mathematical

---

programming. Results demonstrate that state estimation models reported in this study have adequate performance for being used as a data processing tool that determines the most likely operating condition of microgrid systems (i.e., the steady state) for a given available measurement set. The proposed tool is suitable to estimate nodal voltages and other quantities (such as active and reactive powers and power losses) with a reasonable number of iterations. In the case of WLS, the IPOPT solver is able to solve the non-linear problem. However, in the case of WLAV this algorithm could not converge, so the SNOPT solver was used instead, in the Campusgrid test case. For the UKG-16 test microgrid, the Knitro solver was used, as both IPOPT and SNOPT failed to converge. Results suggest that IPOPT is more suitable for solving the WLS case because it is a quadratic problem with non-linear restrictions. In contrast, WLAV requires computing the Hessian because it is a non-linear problem with non-linear restrictions. In this case, it is recommendable to use KNITRO to solve WLAV, because it has fewer restrictions of variables compared to SNPOT. This is of central importance because AMPL modeling language provides automatic generation of first and second derivatives of constraint and objective functions. The WLS model is the most recommendable for this problem, as it has a lower computational cost (four iterations for the Campusgrid-60 test case and three iterations for the UKG-16 test case). In this paper, bad data in measurements and topology influences are not considered for the state estimation. Future work will explore how to incorporate a pre-processing stage to identify measures without gross errors, such as data latency and false data injection attacks, before performing state estimation.

# Chapter 3

## Contribution II

### Design and Implementation of a Machine Learning State Estimation Model for Unobservable Microgrids<sup>1</sup>

**Abstract:** An observable microgrid may become unobservable when sensors are at fault, sensor data is missing, or data has been tampered by malicious agents. In those cases, state estimation cannot be performed using traditional approaches without pseudo-measurements. To address the lack of observability, this article presents the design and implementation of a novel three-phase state estimation method for unobservable and unbalanced AC microgrids, using machine learning techniques, without pseudo-measurements, and under heteroscedastic (i.e., non-constant variance) noise. The proposed machine learning state estimation (MLSE) makes full use of multiple candidate models trained with a small number of power flow simulations via OpenDSS, through random levels of demand and renewable generation in every simulation, enhanced through a proposed Tikhonov regularization operator. To deal with the heteroscedastic nature of measurements, a recursive average model is proposed to accurately estimate the state variables. Results are obtained using real data from a microgrid located on the main campus of the State University of Campinas (UNICAMP), in Brazil. The method can be easily adapted to microgrids with different configurations, distributed energy resources, and measurements. It is shown that the proposed MLSE outperforms the traditional weighted least square (WLS) state estimator.

**Keywords:** Machine learning, microgrids, Moore-Penrose left pseudoinverse, Tikhonov regularization operator, recursive average model.

<sup>1</sup>Contribution II is published as: textbfByron Alejandro Acuña Acurio, Diana Estefanía Chérrez Baragán, Juan Camilo López, Marcos J. Rider and Luiz Carlos Pereira da Silva, "Design and Implementation of a Machine Learning State Estimation Model for Unobservable Microgrids" [40] in IEEE ACCESS 2022, doi: <https://doi.org/10.1109/ACCESS.2022.3224758>

## 3.1 Introduction

State estimation calculates the most-likely values of the state variables in an electrical grid, from real-time available measurements. This information is helpful for the identification of erroneous measurements, on-line parameter estimation, cyber-security assessment, and autonomous energy management systems (EMS) [25]. State estimation has been mostly applied in transmission systems due to their high level of digitization [41]. The direct translation of state estimation techniques from transmission systems to the distribution level is not always possible [42], so that the approaches used in distribution systems and AC microgrids are still challenging [43, 29, 31]. On the other hand, incorporating uncertainty into state estimation models has also become a relevant research topic due to the proliferation of low-cost smart meters in distribution systems as a replacement of phasor measurement units (PMUs). In this case, noise measurements and uncertainty due to malfunction associated with smart meters are generally unknown and difficult to characterize in real-time applications[44].

The main challenges of applying a traditional state estimation in microgrids can be summarized as follows: (1) microgrids are typically unbalanced three-phase AC distribution systems. Hence, there is an increase in the required number of variables and equations, compared to classical approaches; (2) the  $X/R$  ratios in microgrids are significantly low. As a result, fast decoupled state estimators [45] cannot be used, and conventional DC state estimators become irrelevant [39]; (3) limited observability with only a few measurements available[46]; (4) topological changes (i.e., transition from main grid-connected to islanded mode, and vice-versa) can lead to different operating points in a short amount of time [47]; and (5) PMUs are not as ubiquitous as in transmission systems.

In this context, considerable efforts have been made to develop state estimators for distribution networks [48]. However, very few works have addressed the development of state estimators for AC microgrids [49, 50, 51, 52]. Most static estimators are based on the classical weighted least squares (WLS) method [53]. In [54], the WLS method was applied to multiple microgrids, resulting in high computational complexity. Other works have focused on scrutinizing the problem of dynamic state estimation for microgrids [55]. For instance, in [56] a modified extended Kalman filter (EKF) [57] was proposed to estimate the dynamic variables of generator units, as well as the static variables related to the network. In [58], a distributed dynamic state estimator was proposed to estimate the operation of the energy resources and the status (connected or islanded) of the microgrid. In [59], two techniques for state estimation in droop-controlled islanded microgrids were proposed, via an unscented Kalman filter (UKF) [60] and a non-linear particle filter. Microgrids usually comprise high-speed conversion systems associated with extremely fast dynamics, on the order of nano-seconds. Thus, dynamic state estimation would require

expensive resources for very limited applications. On the other hand, estimation of steady-state electricity variables with scarcity of smart meters, communication networks, and computation technology is more reasonable and cost-benefit approach.

The main differences among the aforementioned works lie in the choice of the state variables and how the measurements are considered. By their choice of state variables, state estimation approaches can be broadly classified into two major groups: (1) node voltage estimators and (2) branch current estimators [48]. Both approaches can be formulated in polar or rectangular coordinates. Regarding data, some methods use raw measurements (i.e., originally measured or forecasted values), and others use equivalent voltage and current measurements. [48]. Forecasted data approaches are not adequate for microgrids able to seamlessly disconnect from the grid and operate in an islanded mode in the event of a disturbance. Nevertheless, most state estimation methods for microgrids are mainly based on WLS using analytical formulations to model the system and truncated iterative methods to solve the non-linear set of equations [61].

Traditional state estimation using WLS is widely used in transmission systems [24], and it has been found to be suitable for observable grid-connected microgrids [62]. The non-linear relationship between the state variables and measurements results in a state-dependent Jacobian matrix that has to be updated at each iteration of the state estimation process. Because of its state-dependent Jacobian (and thus gain) matrix, applying this method to microgrids is computationally expensive[63]. Moreover, WLS requires the meter error variance (usually noted as  $\sigma$ ) as an input parameter to perform the estimation.

Classical WLS approaches consider  $\sigma$  as a known constant. This work addresses the possibility of the measurements having different noise levels. The accurate knowledge of noise characteristics in state estimation methods is a prerequisite for designing a high-performance estimator [42]. However, characterizing noise using low-cost smart-meters can be challenging because the sampling rate ranges from minutes to hours [44]. Previous works considered measurements from (PMUs) which typically have a sampling rate of more than 50 samples per second where noise characterization is feasible [41]. Other works have used machine learning only to assist classical state estimators (as in [64, 65]). However, very few works [66, 67] have addressed the direct use of machine learning state estimation (MLSE) for AC microgrids. State estimation can be considered as a regression problem [68]. Therefore, if the noise ( $\sigma$ ) is assumed known, constant, independent, and identical distributed (*iid*) from the parameters (i.e., homoscedastic), it is possible to develop a model for mapping the available measurements to states variables using supervised learning to train any regression model [68] with a suitable dataset. In this work, the noise is considered input-dependent (i.e., heteroscedastic), which is a more realistic assumption when low-cost smart meters with low sampling rates are used instead of PMUs[44]. Heteroscedastic noise encodes the possibility of the measurements to have different variances  $\sigma$  for various real-time observations. On the other hand, the proposed approach estimates

the state without specify (i) an initial guess estimation point (a.k.a., *flat start*), nor (ii) an explicit definition of the measurements' variances ( $\sigma$ ), as is the case of the classic WLS approach.

While many existing methods focus on solving state estimation problems under simplified assumptions, such as homoscedastic measurements, this manuscript argues the importance of exploring ways for merging simulation and machine learning to solve state estimation under varying conditions, such as heteroscedastic measurements. A primary motivation of this study is the potential future applications to handle the unexpected changes in the dynamic and stochastic nature of the microgrids that can be numerically intractable using classical approaches.

The proposed MLSE uses a synaptic matrix of weights  $\mathbf{W}$  that captures the functional relationship between state variables and the available measurements, in order to estimate the state variables accurately. The training model is based on Moore-Penrose left pseudoinverse, enhanced through a Tikhonov regularization. A robust MLSE against heteroscedastic uncertainty is obtained using an approach named in this paper as *recursive average model*. AC power flow simulations via OpenDSS [69] are used as the source of knowledge. Results are obtained using data from a real-world microgrid located at the main campus of the State University of Campinas (UNICAMP), in São Paulo, Brazil. An overview of the microgrid is shown in Fig. 3.1 [1]. The performance and accuracy of the proposed MLSE is analyzed during grid-connected and islanded operation modes with noisy measurements, each with a different signal-to-noise ratio (SNR). In summary, the contributions of this paper are as follows:

A novel MLSE for unobservable three-phase unbalanced AC microgrids, that does not require previous knowledge of the meter error variance  $\sigma$  to deal with the heteroscedastic nature of measurements;

A proposed regularization operator that captures the functional relationship between state variables and available measurements;

A novel learning approach named *recursive average model* for robustness against heteroscedastic uncertainty.

## 3.2 Classical State Estimation

Given a set of  $m$  available noisy measurements  $\mathbf{z} = (z_1, z_2, \dots, z_m)^T$ , state estimation determines the *most-likely*  $n$  state variables  $\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n)^T$  of the microgrid. In this case, the required state variables are the voltage magnitudes ( $V_{i,p}$ ) and phase voltage angles ( $\theta_{i,p}$ ),

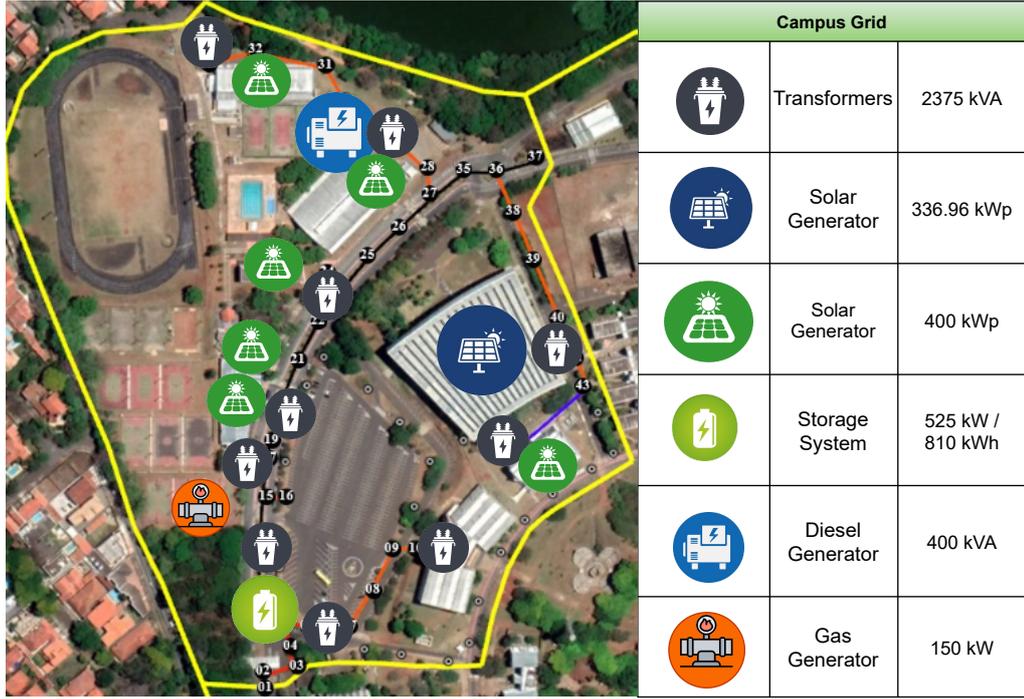


Figure 3.1: Aerial view of the campus microgrid at the State University of Campinas (UNI-CAMP) [1].

where  $i$  is the bus number, given by  $i \in \{1, 2, 3, \dots, n_b\} \equiv \mathbf{B}$ ,  $n_b$  is the total number of buses, and  $p$  is the phase defined by  $p \in \{a, b, c\}$ . Thus, the set of state variables consists of  $3n_b$  elements of the voltage magnitudes, plus  $3(n_b - 1)$  elements of the phase voltage angles. Note that an estimation of  $\mathbf{x}$  is sufficient to determine all remaining electrical variables [61]. Set  $\mathbf{z}$  contains different types of measurements, such as active power injection ( $P_{i,p}$ ), reactive power injection ( $Q_{i,p}$ ), current magnitude injection ( $I_{i,p}$ ), active power flow ( $P_{ij,p}$ ), reactive power flow ( $Q_{ij,p}$ ), current magnitude ( $I_{ij,p}$ ), among others,

where  $j \in \mathbf{B}$  and  $ij$  is a branch. Thus, given a set of measurements  $\mathbf{z}$ , an analytical expression that maps the estimated state variables  $\hat{\mathbf{x}}$  with  $\mathbf{z}$  is given by (3.1).

$$\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix}_{m \times 1} = \begin{bmatrix} h_1(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n) \\ h_2(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n) \\ \vdots \\ h_m(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n) \end{bmatrix}_{m \times 1} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix}_{m \times 1} \quad (3.1)$$

In which  $\hat{\mathbf{x}} \in \mathbb{R}^n$  contains the estimated state variables,  $\mathbf{z} \in \mathbb{R}^m$  contains the available measurements, and  $\mathbf{e} \in \mathbb{R}^m$  is the added noise for each measurement. Note that if the number of real-time measurements  $\mathbf{z}$  is lower than the number of required state variables  $\hat{\mathbf{x}}$ , observability might not be achievable and, as a consequence, it is impossible to calculate state variables without using pseudo-measurements[39]. The minimum amount of measurements  $m_{\min}$  needed for the method to work is  $m_{\min} = 2n - k$ , where  $k$  is the

number of defined slack buses. In this case, the number of state variables is  $2n - k$ , assuming the microgrid contains  $n$  buses, then the microgrid is described by  $2n$  variables, namely  $n$  voltage absolute values and  $n$  voltage angles. However, to improve accuracy, the number of redundant measurements should be higher. In practice, a value of  $m \approx 4n$  is often considered reasonable [39]. Thus, (3.1) can be written as:

$$\mathbf{z} = \mathbf{h}(\hat{\mathbf{x}}) + \mathbf{e} \quad (3.2)$$

where, vector  $\mathbf{h}(\hat{\mathbf{x}})$  are functions that maps the state variables to the measurements and the stochastic noise  $\mathbf{e}$  is modeled as multivariate Gaussian  $\mathbf{e} \sim \mathcal{N}(0, \Sigma_e)$ . The solution of the state estimation problem using WLS approach can be expressed as follows:

$$\min \left\{ \sum_{i=1}^m W_i r_i^2 \right\} \quad (3.3)$$

where  $i = \{1, 2, 3, \dots, m\}$  is the measurement used in the state estimation,  $W_i$  is the weighting factor obtained from the stochastic noise model of each measurements. This is considered to be constant white Gaussian noise with zero mean. Thus,  $W_i = \sigma_i^{-2}$  and  $r_i$  are residuals computed as:

$$r_i = h_i(\hat{\mathbf{x}}) - z_i \quad (3.4)$$

As indicated by equation (3.3), the presence of different noise parameters  $\sigma$  in the objective function implies that dissimilar evaluations of the same set of measurements  $\mathbf{z}$  will lead to erroneous states values  $\hat{\mathbf{x}}$ .

### 3.3 Data Generation

Training a machine learning model with heteroscedastic noise measurements introduces uncertainty in the machine learning model. In order to deal with this challenge, the following data generation process is performed. From the statistical perspective, state variables generate a random vector  $\mathbf{x}$  obtained from random variables that follow a multivariate normal distribution  $\mathcal{X} \sim \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$ . In this case, given  $\mathbf{x}$ , it is possible to obtain a  $m$ -tuple of noiseless measurements  $\mathbf{h}(\hat{\mathbf{x}})$  using well-known analytical or computational microgrid models  $h(\cdot)$ , such as those in [70]. In practice, the information about those parameters that define the multivariate normal distribution  $\mathcal{X}$ , is not always available. However, using samples generated from power flow simulations, it is possible to train the machine learning model  $\mathbf{W}_{\mathcal{D}_{\text{train}}, \lambda_{\text{best}}}$  to capture the non-linear relationship between the state variables  $\mathbf{x}$  and measurements  $\mathbf{z}$ . Subscript  $\mathcal{D}_{\text{train}}, \lambda_{\text{best}}$ , refers to the training dataset and the regularization hyperparameter used to fit the model  $\mathbf{W}$ . This will be explained extensively in the next sections. In this paper, the microgrid was first modeled

in OpenDSS [69] to generate a noiseless dataset<sup>2</sup>  $\mathcal{D}_\infty = \left( \mathbf{z}^{(i)} \in \mathcal{Z}, \mathbf{x}^{(i)} \in \mathcal{X} \right)_{i=1}^\delta$ , using traditional power flow simulations, with random levels of demand and renewable generation in every event  $\delta$ . As shown in Fig. 3.2 one thousand and five hundred events ( $\delta = 1500$ ) were generated.

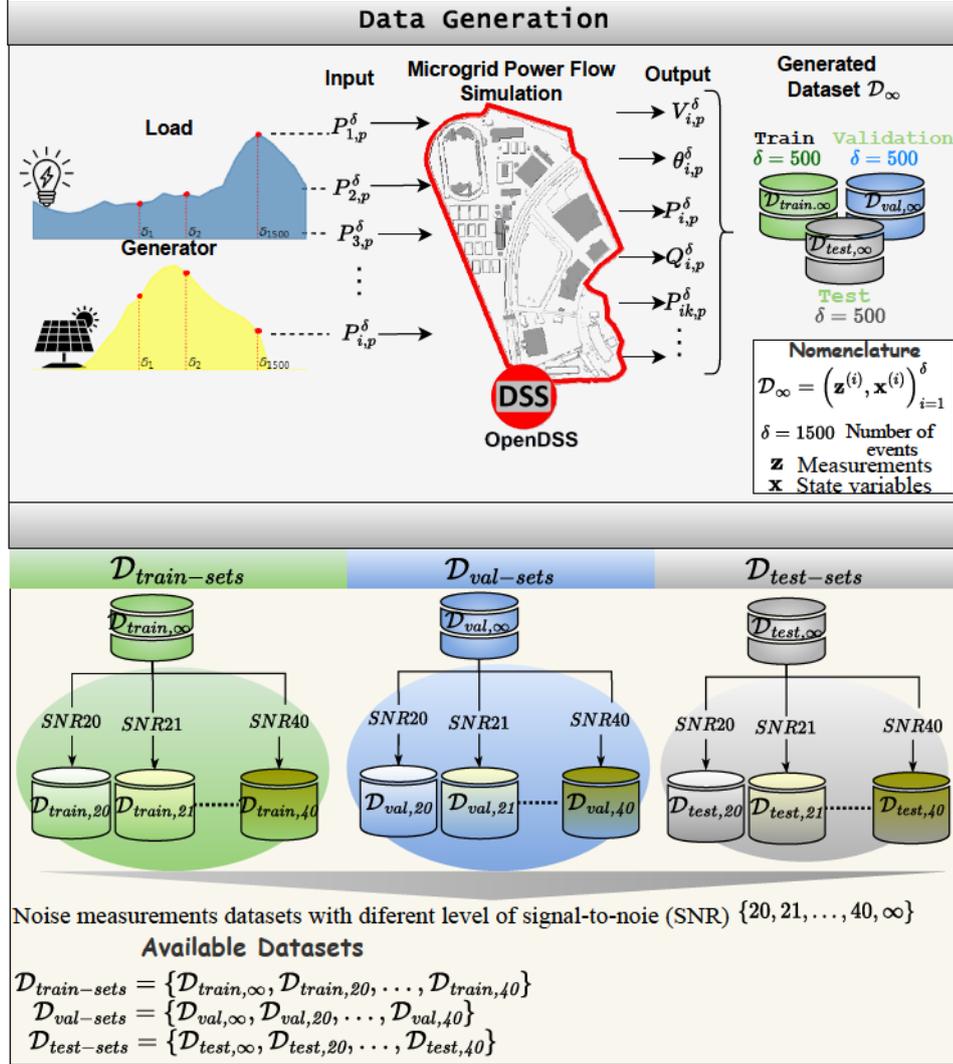


Figure 3.2: Proposed data event generator.

For numerical stability, a *normalizing* process to make all the state variables  $\mathbf{x}$  lie between 0 and 1, in the generated data  $\mathcal{D}_\infty$ .

### 3.3.1 Heteroscedastic noise measurements

The noise used in the proposed MLSE is heteroscedastic. Thus, it can be defined as follows:

$$\mathbf{z} = \mathbf{h}(\hat{\mathbf{x}}) + \mathbf{e}(\mathbf{z}), \quad \mathbf{e}(\mathbf{z}) \sim \mathcal{N}(0, g(\mathbf{z})) \quad (3.5)$$

<sup>2</sup>In this work, notation  $\infty$  indicates noiseless measurements, i.e., signal-to-noise (SNR)  $\rightarrow \infty$ . Notation  $(\cdot)_{i=1}^\delta$ , is used to indicate that  $i$  varies from 1 to  $\delta \in \mathbb{N}$ .

To transform the noiseless dataset  $\mathcal{D}_\infty$  into heteroscedastic noise measurements, a signal-to-noise (SNR) ranging from 20 to 40 dB was defined, with increments of 1 dB, based on previous publications [71, 72, 73, 74, 75, 76] that worked with state estimation over homoscedastic measurements. SNR quantifies the measurement error for each measurement using (3.6).

$$\text{SNR}_{[\text{dB}]} = 10 \log_{10} \frac{[z_m^2]}{\sigma_m^2} \quad (3.6)$$

The error  $e_m \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma_m^2)$  used to pollute each measurement  $z_m$  is injected as in (3.5). Note that the state variables  $\mathbf{x}$ , computed in  $\mathcal{D}_\infty$ , correspond to a unique microgrid operating point despite the fact that the measurements  $\mathbf{z}$  are randomly perturbed, causing an increment of uncertainty in each measurement  $z_m$ . The set  $\mathcal{D}_{\text{set}} = \{\mathcal{D}_\infty, \mathcal{D}_{\text{SNR}20}, \mathcal{D}_{\text{SNR}21}, \dots, \mathcal{D}_{\text{SNR}40}\}$ , that contains several noisy datasets, eliminates the need to specify the nature of the measurement error  $\mathbf{e}(\mathbf{z})$  during the training process. As shown in Fig. 3.2, from the available data  $\mathcal{D}_{\text{set}}$  a stratified sampling strategy is employed to select  $\mathcal{D}_{\text{test}} = 500$  records for test at random [77]. Then, a second stratified sampling method is used to divide the rest of the set into a training set  $\mathcal{D}_{\text{train}} = 500$  and a validation set  $\mathcal{D}_{\text{val}} = 500$ . In this case, the available datasets are: (i)  $\mathcal{D}_{\text{train-set}} = \{\mathcal{D}_{\text{train},\infty}, \mathcal{D}_{\text{train},20}, \dots, \mathcal{D}_{\text{train},40}\}$ , (ii)  $\mathcal{D}_{\text{val-set}} = \{\mathcal{D}_{\text{val},\infty}, \mathcal{D}_{\text{val},20}, \dots, \mathcal{D}_{\text{val},40}\}$ , and (iii)  $\mathcal{D}_{\text{test-set}} = \{\mathcal{D}_{\text{test},\infty}, \mathcal{D}_{\text{test},20}, \dots, \mathcal{D}_{\text{test},40}\}$ . Due to the noisy heteroscedastic nature of the measurements, traditional multi-output linear regression methods without regularization are inadequate because different measurements could return the same state variables. Other well-known approaches that use classical regularization methods, such as the Ridge, Lasso & Elastic Net Regression, were tested using Scikit-learn [78] Python library without achieving good results. Therefore, the next section in this paper is concerned with the computation of a machine learning model  $\mathbf{W}_{\mathcal{D}_{\text{train}}, \lambda_{\text{best}}}$  able to perform an approximate solution to obtain estimate state variables from noisy heteroscedastic measurements in unobservable microgrids, since there is no analytical equation to solve this problem.

### 3.4 Proposed MLSE and Regularization Process

The analysis explained for  $\mathcal{D}_{\text{train}}$  and  $\mathcal{D}_{\text{val}}$  will be the same for all the available datasets, for each SNR level at a time. Then, as shown in Fig. 3.2, a tuple of training and validation datasets ( $\mathcal{D}_{\text{train}}, \mathcal{D}_{\text{val}}$ ) is selected for a determined SNR level. In general, a machine learning problem can be solved using the following three steps [79]:

1. Define the machine learning model, for instance single layer networks, feed-forward neural network, recurrent neural networks, etc.
2. Define a metric to evaluate the learning process, like mean squared error (MSE), cross entropy, etc.

3. Define an optimization algorithm to improve the machine learning model such as least squares, gradient descent, etc.

### 3.4.1 Machine Learning Model

Different from classical WLS approaches that perform an optimization process iteratively. The proposed MLSE is based in a geometric point-of-view, in which machine learning is used to fit a parametric model matrix  $\mathbf{W}$  that projects the information of available measurements to state variables. It is possible to use the least square method to find a projection operator  $\mathbf{W}_{\mathcal{D}_{\text{train}}, \lambda_{\text{best}}}$ , which is a matrix whose dimensions depend on the number of  $n$  state variables and  $m + 1$  measurements, as follows:

$$\mathbf{W}_{\mathcal{D}_{\text{train}}, \lambda_{\text{best}}} = \begin{bmatrix} w_{1,0} & w_{1,1} & w_{1,2} & \dots & w_{1,m} \\ w_{2,0} & w_{2,1} & w_{2,2} & \dots & w_{2,m} \\ w_{3,0} & w_{3,1} & w_{3,2} & \dots & w_{3,m} \\ \vdots & & & & \\ w_{n,0} & w_{n,1} & w_{n,2} & \dots & w_{n,m} \end{bmatrix} \quad (3.7)$$

where, the regression parameters of the model  $w_{n,m} \in \mathbb{R}$ , are not known, and must be estimated from the a training dataset  $\mathcal{D}_{\text{train}}$ . In this case,  $n$  by  $m + 1$  regression parameters are used to estimate  $n$  state variable, from  $m$  available measurements, in which the regression parameters  $(w_{i,0})_{i=1}^n$  correspond to the fixed offset for a basis function, a.k.a., *intercept*[79]. In this work, the machine learning model  $\mathbf{W}_{\mathcal{D}_{\text{train}}, \lambda_{\text{best}}}$  was defined as a single layer network, since it has no memory, i.e., the estimated state variables depend only on current measurements. In practice,  $\mathbf{W}_{\mathcal{D}_{\text{train}}, \lambda_{\text{best}}}$  performs an estimation of the state variables from available measurements, using equation (3.8):

$$\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \hat{x}_3 \\ \vdots \\ \hat{x}_n \end{bmatrix} = \begin{bmatrix} w_{1,0} & w_{1,1} & w_{1,2} & \dots & w_{1,m} \\ w_{2,0} & w_{2,1} & w_{2,2} & \dots & w_{2,m} \\ w_{3,0} & w_{3,1} & w_{3,2} & \dots & w_{3,m} \\ \vdots & & & & \\ w_{n,0} & w_{n,1} & w_{n,2} & \dots & w_{n,m} \end{bmatrix} \begin{bmatrix} 1 \\ z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix} \quad (3.8)$$

### 3.4.2 Least-Square Optimization Model

In state estimation, it is required that estimated state variables  $\hat{\mathbf{x}}$  to be as close as possible to the real values of  $\mathbf{x}$ , for all feasible scenarios of microgrid operation, i.e.,  $\forall \delta$ , where  $\delta$  is the number of available events in the training dataset  $\mathcal{D}_{\text{train}}$ . To do so, the well-regarded mean squared error (MSE) [80] is used as the metric to train and validate

the proposed model:

$$\mathbf{MSE}_{\hat{\mathbf{x}}}(\mathbf{x}^{(\delta)}, \hat{\mathbf{x}}^{(\delta)}) = \left[ \|\mathbf{x}^{(\delta)} - \hat{\mathbf{x}}^{(\delta)}\|_2^2 \right] \quad (3.9)$$

where  $[\cdot]$  denotes the expected value and  $\|\cdot\|_2^2$  is the squared Frobenius norm. Since the **MSE** can be assessed as a measure of the aggregated contributions of all estimated variables, it is possible to rewrite (3.9) as:

$$\mathbf{MSE}_{\hat{\mathbf{x}}}(\mathbf{x}^{(\delta)}, \hat{\mathbf{x}}^{(\delta)}) = \frac{1}{n} \sum_{n=1}^n \left[ \mathbf{MSE}_{\hat{x}_n}(x_n^{(\delta)}, \hat{x}_n^{(\delta)}) \right]^2 \quad (3.10)$$

Given that the number of events using  $\delta$  AC power flow simulations is a finite set, the expected value can be approximated by the sample mean computed as:

$$\mathbf{MSE}_{\hat{x}_n}(x_n^{(\delta)}, \hat{x}_n^{(\delta)}) = \frac{1}{\delta} \sum_{j=1}^{\delta} (x_n^{(j)} - \hat{x}_n^{(j)})^2 \quad (3.11)$$

Which is the same metric used to evaluate linear least-squares models in estimation theory and in linear regression models [81]. From equation (3.11), a natural method to determine the best estimation of  $\hat{\mathbf{x}}$  is to obtain the matrix  $\mathbf{W}_{\mathcal{D}_{\text{train}}, \lambda_{\text{best}}}$ , which minimizes the **MSE** applied to the difference between the true state variables  $\mathbf{x}$  and the estimated counterparts  $\hat{\mathbf{x}}$ :

$$\mathbf{W}_{\mathcal{D}_{\text{train}}, \lambda_{\text{best}}} = \arg \min \mathbf{MSE}_{\hat{\mathbf{x}}}(\mathbf{x}^{(\delta)}, \hat{\mathbf{x}}^{(\delta)}) \equiv \arg \min \sum_{n=1}^n \sum_{j=1}^{\delta} (x_n^{(j)} - \hat{x}_n^{(j)})^2 \quad (3.12)$$

Using (3.8) and dataset  $\mathcal{D}_{\text{train}}$ , equation (3.13) is obtained.

$$\begin{bmatrix} \hat{x}_1^0 & \hat{x}_2^0 & \hat{x}_3^0 & \dots & \hat{x}_n^0 \\ \hat{x}_1^1 & \hat{x}_2^1 & \hat{x}_3^1 & \dots & \hat{x}_n^1 \\ \hat{x}_1^2 & \hat{x}_2^2 & \hat{x}_3^2 & \dots & \hat{x}_n^2 \\ \hat{x}_1^3 & \hat{x}_2^3 & \hat{x}_3^3 & \dots & \hat{x}_n^3 \\ \vdots & & & & \\ \hat{x}_1^\delta & \hat{x}_2^\delta & \hat{x}_3^\delta & \dots & \hat{x}_n^\delta \end{bmatrix} = \begin{bmatrix} 1 & z_1^0 & z_2^0 & \dots & z_m^0 \\ 1 & z_1^1 & z_2^1 & \dots & z_m^1 \\ 1 & z_1^2 & z_2^2 & \dots & z_m^2 \\ 1 & z_1^3 & z_2^3 & \dots & z_m^3 \\ \vdots & & & & \\ 1 & z_1^\delta & z_2^\delta & \dots & z_m^\delta \end{bmatrix} \mathbf{W}_{\mathcal{D}_{\text{train}}, \lambda_{\text{best}}}^T \quad (3.13)$$

Throughout this paper, superscript  $T$  denotes transposition, hence, equation (3.13) can be summarized as follows:

$$\hat{\mathbf{X}} = \Phi \mathbf{W}_{\mathcal{D}_{\text{train}}, \lambda_{\text{best}}}^T \quad (3.14)$$

Replacing (3.14) in (3.12), one obtains:

$$\mathbf{W}_{\mathcal{D}_{\text{train}}, \lambda_{\text{best}}} = \arg \min \left\{ \text{tr} \left[ \left( \mathbf{X} - \Phi \mathbf{W}_{\mathcal{D}_{\text{train}}, \lambda_{\text{best}}}^T \right)^T \left( \mathbf{X} - \Phi \mathbf{W}_{\mathcal{D}_{\text{train}}, \lambda_{\text{best}}}^T \right) \right] \right\} \quad (3.15)$$

The above statement can be solved as:

$$\frac{\partial}{\partial \mathbf{W}_{\mathcal{D}_{\text{train}}, \lambda_{\text{best}}}} \left\{ \text{tr} \left[ \left( \mathbf{X} - \Phi \mathbf{W}_{\mathcal{D}_{\text{train}}, \lambda_{\text{best}}}^T \right)^T \left( \mathbf{X} - \Phi \mathbf{W}_{\mathcal{D}_{\text{train}}, \lambda_{\text{best}}}^T \right) \right] \right\} = 0 \quad (3.16)$$

A possible closed-form solution known as a Moore-Penrose left pseudo-inverse [82] for (3.16) is shown in (3.17). However, this is not a feasible solution since the Gramian matrix  $(\Phi^T \Phi)^{-1}$  is singular without a regularization process, as explained in the next section.

$$\mathbf{W}_{\mathcal{D}_{\text{train}}}^T = (\Phi^T \Phi)^{-1} \Phi^T \mathbf{X} \quad (3.17)$$

### 3.4.2.1 Application of the Tikhonov Regularization

In practice, the linear system of equations in (3.13) is an *ill-posed* problem because it does not meet the following three Hadamard criteria for well-posedness: (i) for all admissible data, a solution exists, (ii) for all admissible data, the solution is unique, and (iii) the solution depends continuously on the data [83]. Thus, small perturbations in measurements generate large errors in  $\hat{\mathbf{x}}$ , when the Moore-Penrose left pseudoinverse is used (i.e., equation (3.17)) to fit the parametric model  $\mathbf{W}_{\mathcal{D}_{\text{train}}}$  without regularization. This solution is useless due to the severe propagation of large errors, caused by large norm of  $\mathbf{W}_{\mathcal{D}_{\text{train}}}$  matrix. A matrix with a large norm is called an ill-conditioned matrix. An ill-conditioned matrix can take a unit-length vector and stretch it by a large amount. Thus small uncertainties in the domain vector get magnified and lead to large uncertainties in the range. Linear systems of equations, such as (3.13), with a matrix of the weights  $\mathbf{W}_{\mathcal{D}_{\text{train}}}$ , are often referred as linear discrete ill-posed problems [84]. The standard way to obtain stable solutions is to modify the problem by replacing (3.17) with a nearby problem, whose solution is less sensitive to large errors. This replacement is known as regularization [85, 86, 87]. While many methods focus on use of *zeroth order Tikhonov* without considering  $w_{n,0}$  [88], such as Ridge regression model, this work proposes a new regularization operator  $\mathbf{R}$  that adds terms containing regularization parameters to (3.17) in order to control the norm of  $\mathbf{W}_{\mathcal{D}_{\text{train}}}$  and obtain an stable estimation, as shown in Fig. 3.3.

The remainder of this section describes the regularization process shown in Fig. 3.3, which details the process to select the best regularization hyperparameter  $\lambda$ . This is of central importance for equation (3.18) instead of the not regularized least squares solution, in equation (3.17), aiming at properly characterizing  $\mathbf{W}_{\mathcal{D}_{\text{train}}, \lambda}$ , for a given set of noisy measurements.

To the best of the authors' knowledge, the regularization operator  $\mathbf{R}$  presented in this work has yet to be proposed. Therefore, the solution is obtained by a closed-form

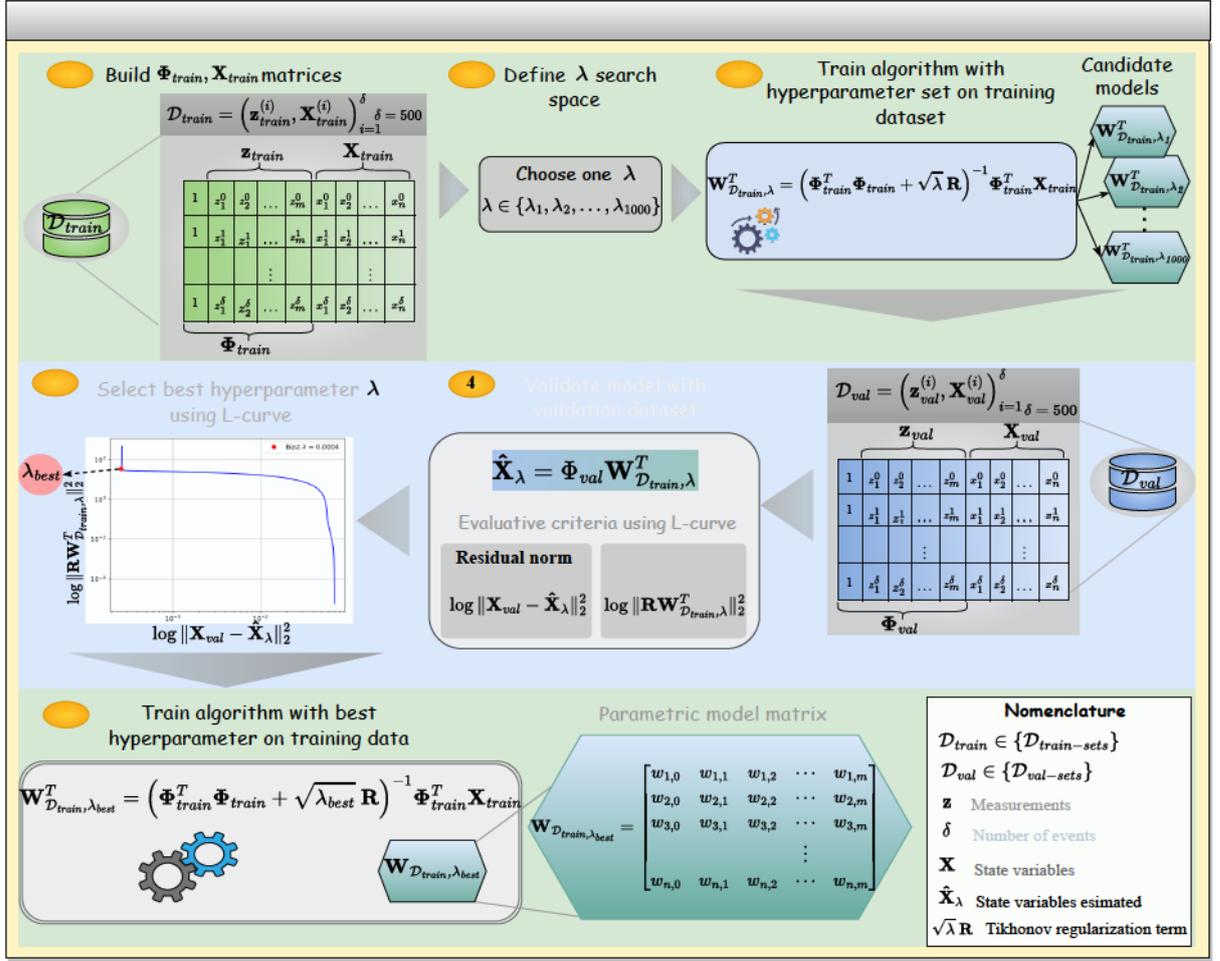


Figure 3.3: Proposed MLSE regularization process.

solution, similar to the one based on the pseudo-inverse, given by:

$$\mathbf{W}_{\mathcal{D}_{train}, \lambda}^T = (\Phi^T \Phi + \sqrt{\lambda} \mathbf{R})^{-1} \Phi^T \mathbf{X} \quad (3.18)$$

The design of the regularization is governed by available knowledge or model assumptions about the state estimation (i.e., prior information) and, to lesser extent, by the implementation or computational complexity. To solve linear discrete ill-posed problems  $\lambda$  in equation (3.18) is known as the Tikhonov regularization hyperparameter, and  $\mathbf{R}$  is known as the regularization operator [89]. The idea is to shift the spectrum of  $\Phi^T \Phi$ , i.e., shift the singular values away from zero. A singular value is the positive square root of an eigenvalue of the symmetric matrix  $\Phi^T \Phi$ . Note that this expression  $(\Phi^T \Phi)^{-1}$  in equation (3.17) is invertible if and only if all the singular values of  $\Phi^T \Phi$  matrix are non-zero. It is worth mentioning that  $\lambda > 0$  because it serves as the coefficient that shifts the diagonals of the  $\Phi^T \Phi$  matrix, a.k.a, singular moment matrix [90]. Parameter  $\lambda$  provides balance between the data fidelity (first term:  $\Phi^T \Phi$ ) and prior information assumptions (the remaining terms  $\mathbf{R}$ ). Note that, with a  $\lambda = 0$ , equation (3.18) is equal to (3.17), making the term  $(\Phi^T \Phi)^{-1}$  singular again.

The proposed MLSE regularization process is shown in Fig. 3.3. A suitable value of  $\lambda$  generally is not explicitly known and must be determined using the validation dataset  $\mathcal{D}_{\text{val}}$  during the training process. To do so, a  $\lambda$  search space with dissimilar values must be defined as  $\lambda \in \{\lambda_1, \lambda_2, \dots, \lambda_{1000}\}$ . In this work, equation (3.18) is used to train candidate MLSE models with different  $\lambda$ 's, creating the following model search space:  $\mathbf{W}_{\mathcal{D}_{\text{train}}, \lambda} \in \{\mathbf{W}_{\mathcal{D}_{\text{train}}, \lambda_1}, \mathbf{W}_{\mathcal{D}_{\text{train}}, \lambda_2}, \dots, \mathbf{W}_{\mathcal{D}_{\text{train}}, \lambda_{1000}}\}$ . To determine the best MLSE model  $\mathbf{W}_{\mathcal{D}_{\text{train}}, \lambda_{\text{best}}}$ , obtained with a specific Tikhonov regularization hyperparameter  $\lambda_{\text{best}}$ , the *L-curve search methodology* is performed [91]. The *L-curve* is a log-log plot of the norm of a regularized solution  $\|\mathbf{R}\mathbf{W}_{\lambda}^T\|_2^2$  versus the norm of the corresponding residual norm  $\|\mathbf{x}^{(\delta)} - \hat{\mathbf{x}}^{(\delta)}\|_2^2$ . It is a convenient graphical tool for characterizing the trade-off between the size of regularized solutions and their fit to the training dataset, as the hyperparameter  $\lambda$  changes. Graphically, the best  $\lambda$  is the one located in the corner of the *L-curve*, as shown in Fig. 3.3. In this work, it was empirically determined that the regularization operator  $\mathbf{R}$  in equation (3.19) shows a stable solution for the non-linear three-phase state estimation problem in unbalanced AC microgrids.

$$\mathbf{R} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & -1 & 1 \end{bmatrix} \quad (3.19)$$

where,  $\mathbf{R} \in \mathbb{R}^{(m+1) \times (m+1)}$ .

The regularization operator  $\mathbf{R}$  is inspired by the work in [92]. In this case,  $\mathbf{R}$  plays the role of a penalizing filter, it is done by considering the anti-reflective and high-order boundary conditions [93]. Both these conditions were introduced in [94] and [95], respectively. In this case, the boundary conditions of  $\mathbf{R}$  is determined by the first and last rows based on the first-order Tikhonov regularization [96], which penalizes deviations from a constant model, i.e., it favors “flat” (constant) solutions and penalizes gradients, working as a first-difference operator in the boundary conditions. On the other hand, the interior rows (rows 2 to  $m$ ) of  $\mathbf{R}$  penalize model “roughness” or bumpiness (curvature) rather than model gradient, based on the second-order Tikhonov regularization [96]. This favors “smooth” (constant gradient) solutions, working as a second-difference operator.

### 3.4.3 Recursive Average Model

As is shown in Fig. 3.2, instead of using multiple  $\mathbf{W}_{\text{SNR},\lambda_{\text{best}}}$ , one for each SNR level, a novel recursive average model is proposed to generate a single feasible model  $\tilde{\mathbf{W}}$  for the complete SNR interval of noisy measurements, as follows:

1. Train a model  $\mathbf{W}_{\text{SNR},\lambda_{\text{best}}}$  using the MLSE regularization process of Fig. 3.3 with each pair of available training and validation datasets for different SNR levels, obtained as in Fig. 3.2.
2. Use equation (3.20) to average the models recursively, as explained in Fig. 3.4

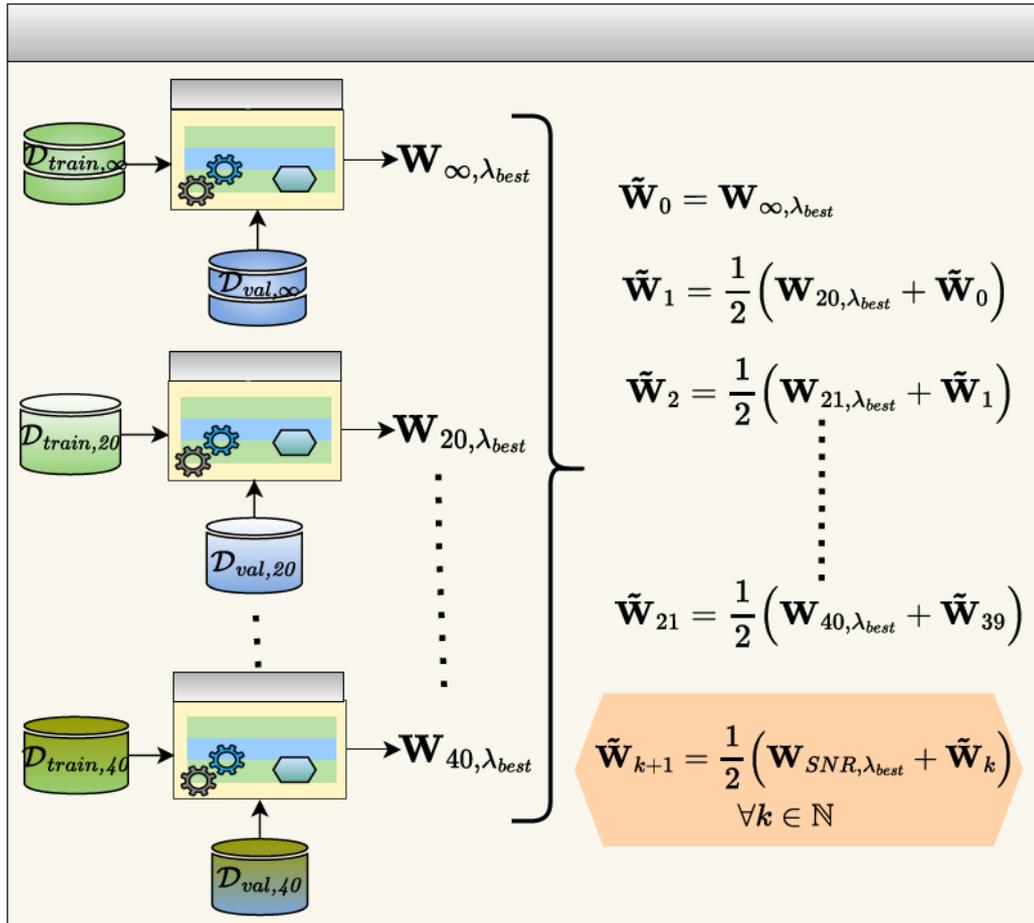


Figure 3.4: Recursive Average Model.

$$\tilde{\mathbf{W}}_{k+1} = \frac{1}{2} \left( \mathbf{W}_{\text{SNR},\lambda_{\text{best}}} + \tilde{\mathbf{W}}_k \right) \quad \forall k \in \mathbb{N} \quad (3.20)$$

## 3.5 Materials and Methods

This section first describes the microgrid system used as a case study. Then, section 3.5.1 explains the mathematical background of hypothesis tests and the criteria used for

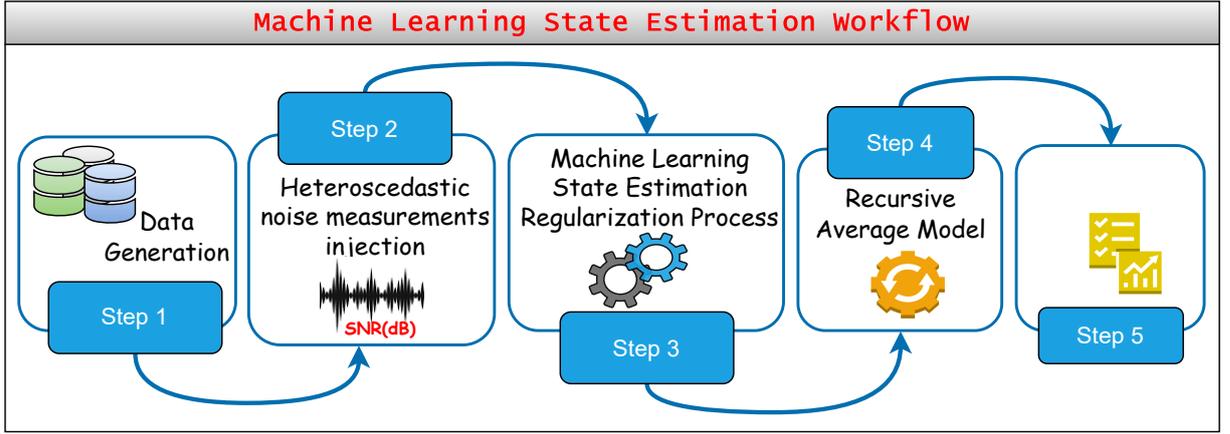


Figure 3.5: Workflow of the proposed MLSE.

validation purposes of the proposed MLSE model. Numerical results obtained using all different scenarios are presented and comprehensively discussed in section 3.6.

A real-world AC microgrid located at the main campus of the State University of Campinas (UNICAMP), in São Paulo, Brazil [1], with 320 state variables ( $V_{i,p}$  and  $\theta_{i,p}$ ) and 198 measurements ( $P_{i,p}$ ,  $Q_{i,p}$ ,  $I_{i,p}$ ) was modeled using OpenDSS [69], and the proposed MLSE via the workflow in Fig. 3.5 programmed in Python 3.8 [97]. From the WLS perspective this case of study is unobservable, without using pseudo-measurements.

### 3.5.1 Statistical analysis

To analyze the accuracy of the proposed MLSE, a box-plot diagram is computed using the available  $\mathcal{D}_{\text{test-set}}$  for each SNR, as is shown in Fig. 3.2. The results of the estimation are compared against the actual power flow calculations using (3.9). In this case, two statistical tests are performed. First, a hypothesis test on the mean value is performed with the purpose of evaluating whether the estimated state variables on each SNR level are statistically equal to the reference given by the power flow simulations. In the second test, known as the homogeneity test, each state variable  $\mathbf{x}$  is analyzed separately with the intention of evaluating whether the set of estimated values  $\hat{\mathbf{x}}$  (using different noise levels) has the same distribution.

**Test of the mean:** According to [98], a hypothesis is *a statement about the parameters of one or more populations*. In this case, one hypothesis should indicate that there is no difference between the estimated state variables and the references values of the test dataset, while the other hypothesis indicates that they are different, i.e.:

$$\mathcal{H}_0: \hat{\mathbf{x}} = \mathbf{x},$$

$$\mathcal{H}_1: \hat{\mathbf{x}} \neq \mathbf{x},$$

where  $\mathcal{H}_0$  is the null hypothesis and  $\mathcal{H}_1$  is the alternative hypothesis.

To test whether the null hypothesis is true or not, the  $t$ -score was chosen as the test statistic since the standard deviation of the population is unknown. A critical region was computed, following the procedure described in [98], with a significance level of 0.05 (the probability of rejecting the  $\mathcal{H}_0$  when it is true).  $\mathcal{H}_0$  is rejected when  $\hat{\mathbf{x}}$  lies outside the critical region. Otherwise, the test fails to reject the null hypothesis, and one could conclude that  $\hat{\mathbf{x}}$  and  $\mathbf{x}$  are the same.

**Homogeneity test:** On the other hand, the homogeneity test evaluates  $n$  populations of interest, divided into  $k$  categories. The chi-squared method was performed by processing the residuals to find out whether the estimated state variables  $\hat{\mathbf{x}}$  contain errors.

$$\chi^2 = \sum_{i=1}^n \frac{(\hat{x}_i - x_i)^2}{x_i} \quad (3.21)$$

The hypotheses were the following:

$\mathcal{H}_0$ : The populations of  $\hat{\mathbf{x}}$  are homogeneous, i.e., the set of estimated state variables do not contain errors.

$\mathcal{H}_1$ : The populations of  $\hat{\mathbf{x}}$  are not homogeneous, i.e., the set of estimated state variables contains errors.

The decision is made considering the  $\chi_{\text{ref}}^2$  statistic with  $c = (n - 1)(k - 1)$  degrees of freedom. In this case, the test was run for each level of noise. This is, with 20 populations defined by the estimated state variables with different values of SNR, each with two categories (voltage magnitude or voltage angle) and a significance level of  $\alpha = 0.01$ .  $\chi_{\text{ref}}^2 = 37.566$ , and the test fails to reject the null hypothesis for computed  $\chi^2$  lower than  $\chi_{\text{ref}}^2$ . In this case, one could conclude that there are no errors in the estimated state variables. The critical region and  $\chi^2$  were computed using the hypothesis test functions of Python software [97].

## 3.6 Numerical Results

In all experiments<sup>3</sup>, the basic metric used to assess the performance of the MLSE is the mean-squared error (**MSE**) between the true value of the state variables  $\mathbf{x}$  (i.e., *ground truth*) and the output offered by the MLSE model ( $\hat{\mathbf{x}}$ ). As shown in Fig. 3.6, the best  $\lambda$  was 0.0004, using a search space that ranges 1e-10 to 1e+10.

The results using the  $\mathcal{D}_{\text{test-set}}$  and the Machine Learning State Estimation (MLSE) shown in Fig. 3.7 are summarized in Fig. 3.8 to Fig. 3.11. These figures are organized as follows: Boxplots show the **MSE** for all the estimated state variables, using the proposed

<sup>3</sup>All tests were performed using a workstation with an Intel(R) Core(TM) i7 870 processor and 8 GB RAM. In this case the average elapsed time for estimating the state variables was 0.000054s.

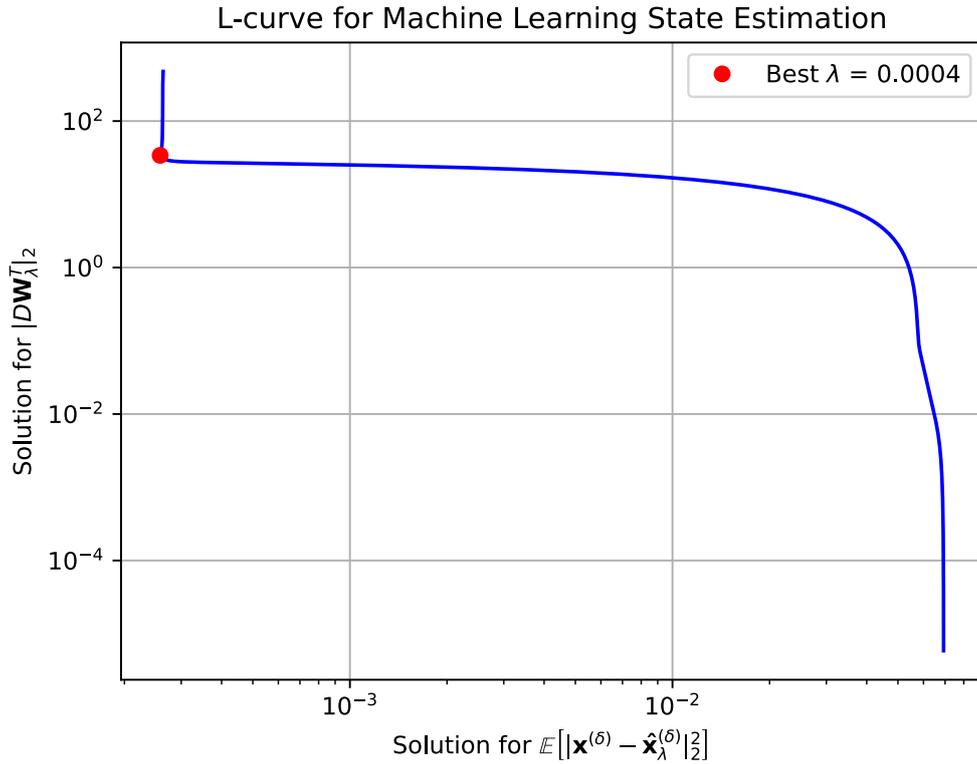


Figure 3.6: L-curve for the proposed MLSE.  $\lambda$  search space from  $1e-10$  to  $1e+10$ .

MLSE model. The boxplot with label “ $\infty$ ” shows the performance of the MLSE model with noiseless measurements.

In this case, the best result using noiseless measurements  $\mathbf{z}$  was approximately  $1e-5$ , as shown in Figs. 3.8 to 3.11. The worst result was  $1e-4$ , approximately. These results suggest that the MLSE model is adequate for heteroscedastic noisy measurements. This is in contrast with traditional WLS state estimation approaches previously tested in the University of Campinas, that reported about  $1e-3$ , using homoscedastic measurements and pseudo-measurements to deal with lack of observability(cf. [99] for a broader discussion on this).

Regarding testing of the mean, the  $t$ -values of the test are presented on Figs. 3.12 - 3.13 (tops). It can be seen that all  $t$ -values are close to zero. This means that there are no  $t$ -values within the rejection regions determined by the critical  $t$ -value= $\pm 1.974$ . Instead, all the estimated state variables lay inside the critical region  $-1.974 < t\text{-value} < 1.974$ . This shoes that the reference state variables and the estimated state variables are equivalent. Note that, since it is known that the  $t$ -distribution has zero mean, the  $t$ -values can be either positive or negative.

In addition, the  $p$ -values are shown in Figs. 3.12 - 3.13 (middles). They quantify the probability of observing a more extreme test statistic, in the direction of the alternative hypothesis. In a nutshell, if all  $p$ -values are larger than the chosen threshold (5%), this would indicate that the observation cannot occur by mere chance, and the null hypothesis

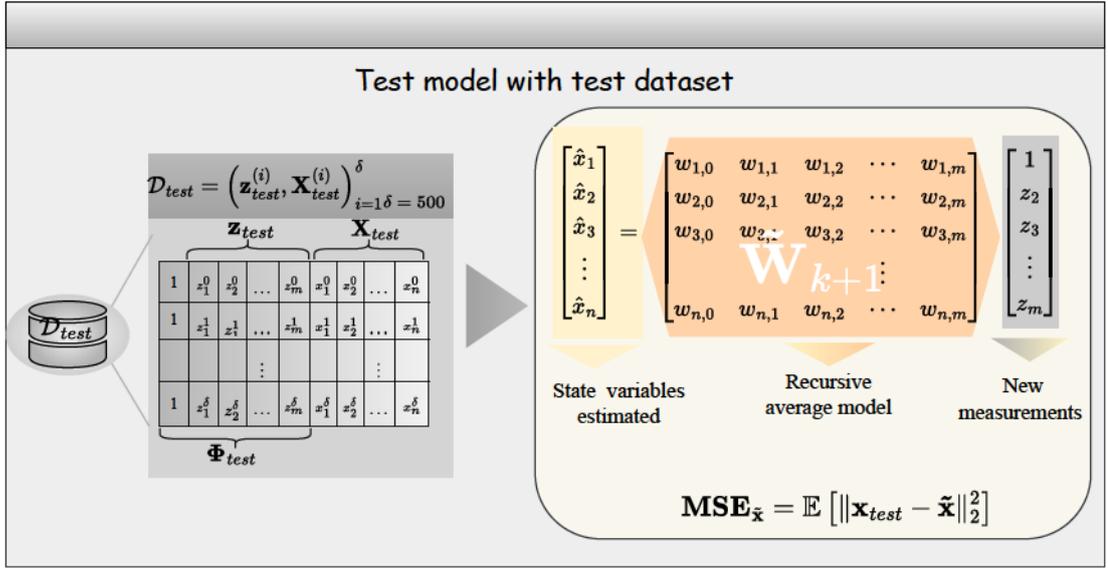


Figure 3.7: Test Model.

of equal population means would not be rejected. In this case, all  $p$ -values were larger than the threshold. Consequently, the test fails to reject the null hypothesis i.e. all the estimated state variables  $\hat{\mathbf{x}}$  should not be considered different than the *ground truth*  $\mathbf{x}$  on any  $V_{i,p}$  or  $\theta_{i,p}$ .

Finally, the distribution of the  $p$ -values is presented in Figs. 3.12 - 3.13 (bottoms), to show how much the test correctly concludes that the estimated state variables are equivalent to the reference state variables. In this work, the deployed test of the mean concludes that estimated and reference state variables are equivalent, with a confidence level of 95%.

In order to test whether the estimated state variables and the references have the same distribution, the homogeneity test explained in Section 3.5.1 was deployed. In this case, all tests return  $p$ -values greater than the  $\alpha = 0.01$ , as shown in Fig. 3.14. Thus, with a level of significance of 1%, there is no evidence to conclude that the distribution of the reference variables is different than the distribution of the estimated state variables.

### 3.6.1 Computational complexity analysis

In order to have a performance value of the proposed MLSE algorithm on running time, the asymptotic execution time as a fundamental measure of the computational complexity efficiency is used. The asymptotic analysis makes it possible to determine the complexity of the asymptotic time. In this paper, the asymptotic notations  $\Theta, O, \Omega$  are used to describe the worst-case time complexity of the MLSE. To be consistent with the notation in equation (3.8) and Fig. 3.7, the entry corresponding to the  $i$ -th row and  $j$ -th column is denoted as  $\mathbf{W}[i, j]$ . Similarly, the  $i$ -th entry in the measurements vector  $\mathbf{z}$  is denoted as  $\mathbf{z}[i]$ . Thus, the vector of estimated state variables  $\hat{\mathbf{x}}$  of length  $n$  and its  $i$ -th

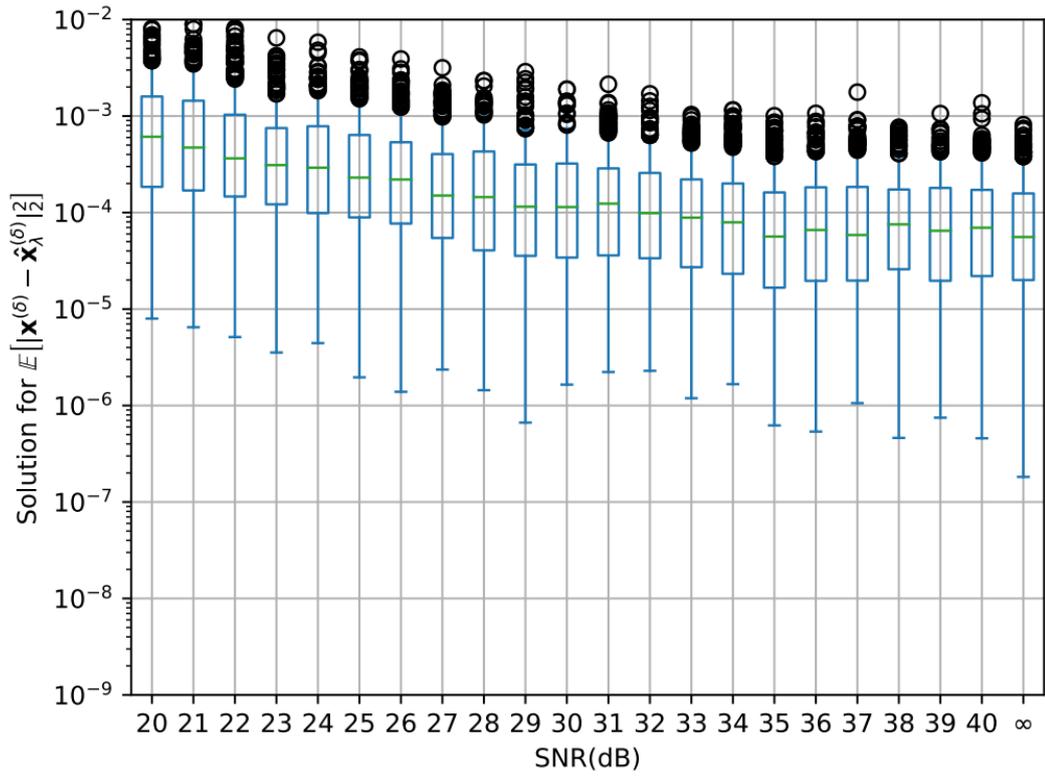


Figure 3.8: Results in terms of mean squared error of all Voltage magnitudes  $V_{i,p}$  in grid-connected mode.

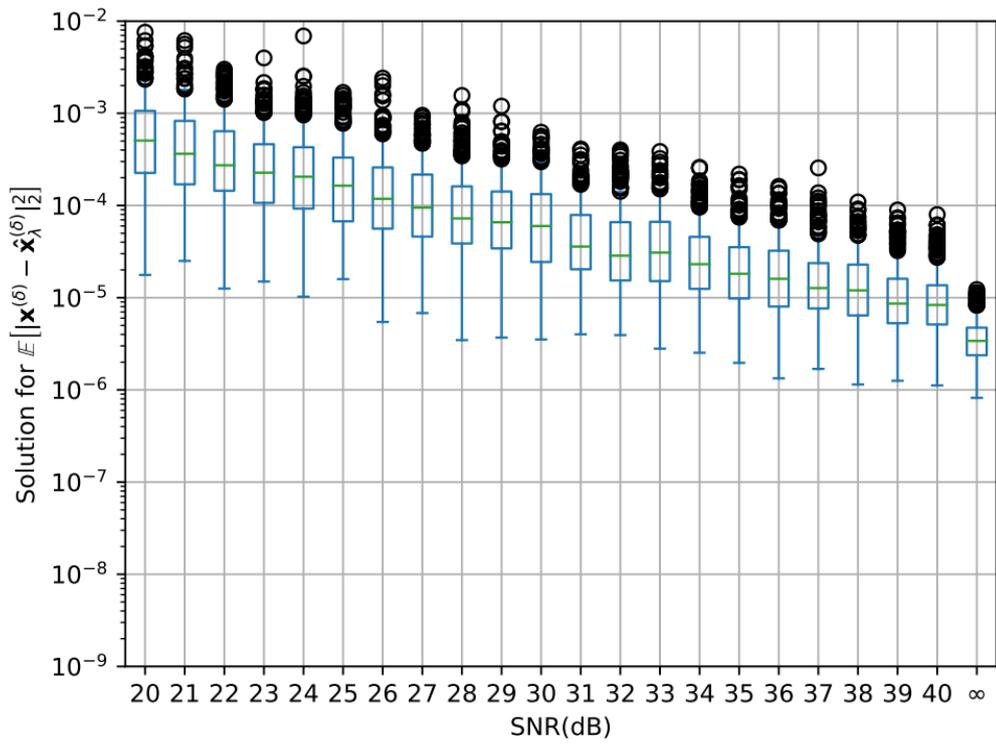


Figure 3.9: Results in terms of mean squared error of all Voltage phases  $\theta_{i,p}$  in grid-connected mode.

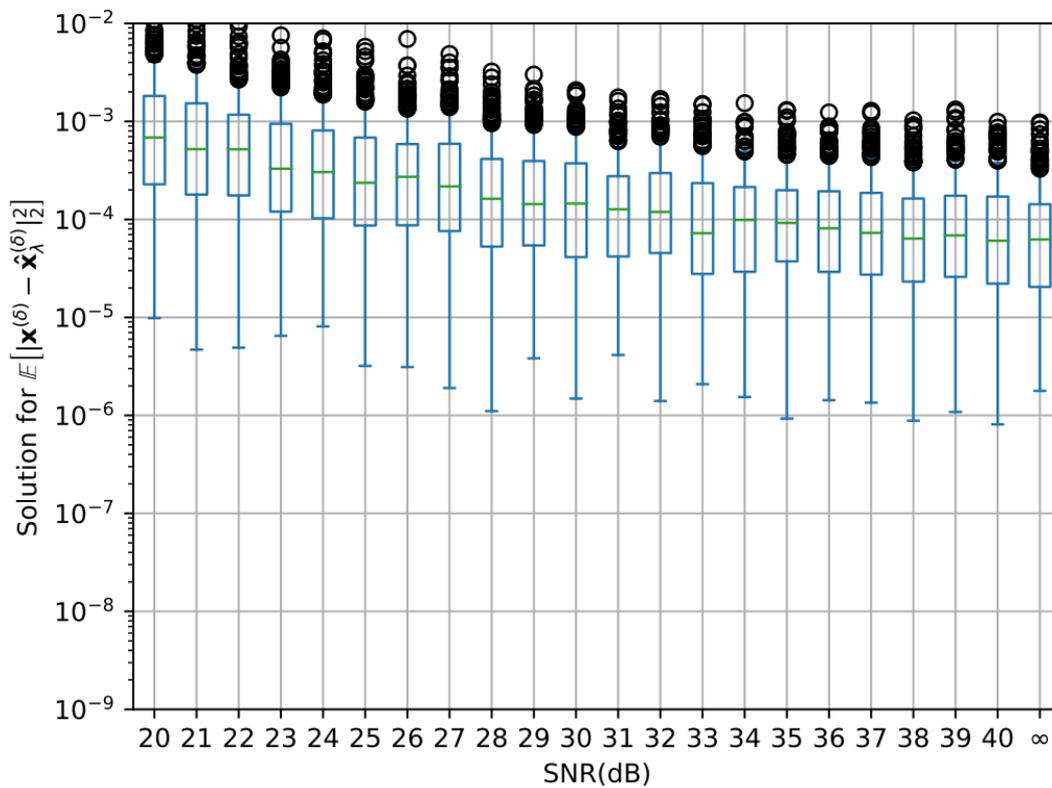


Figure 3.10: Results in terms of mean squared error of all Voltage magnitudes  $V_{i,p}$  in islanded mode.

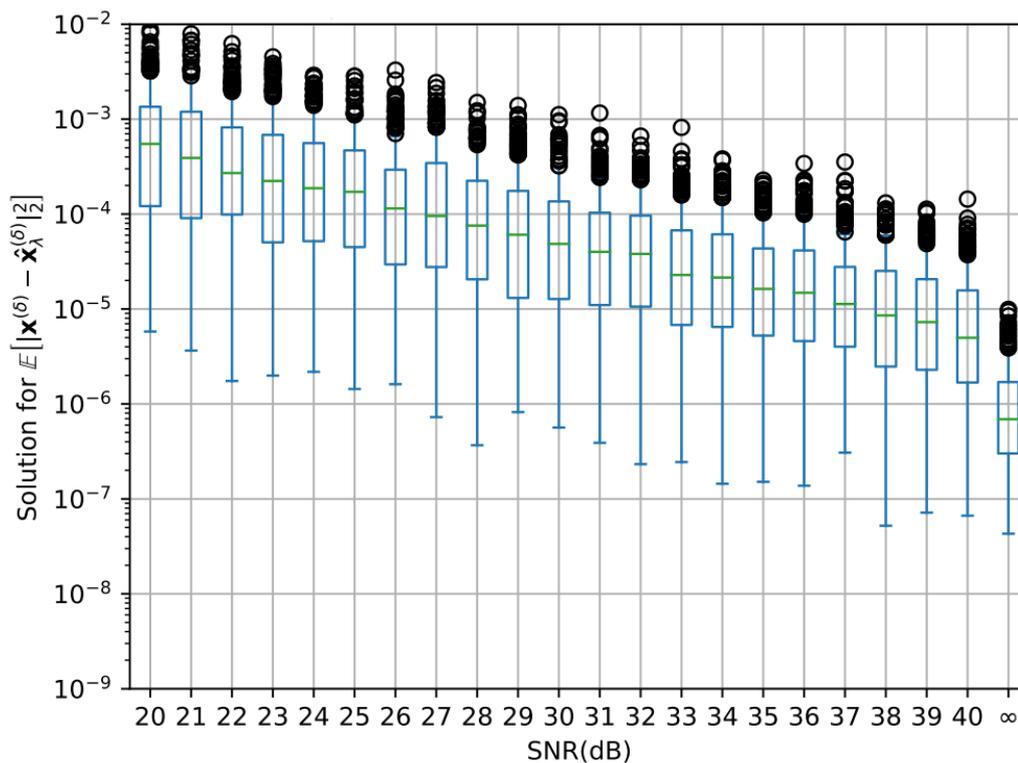
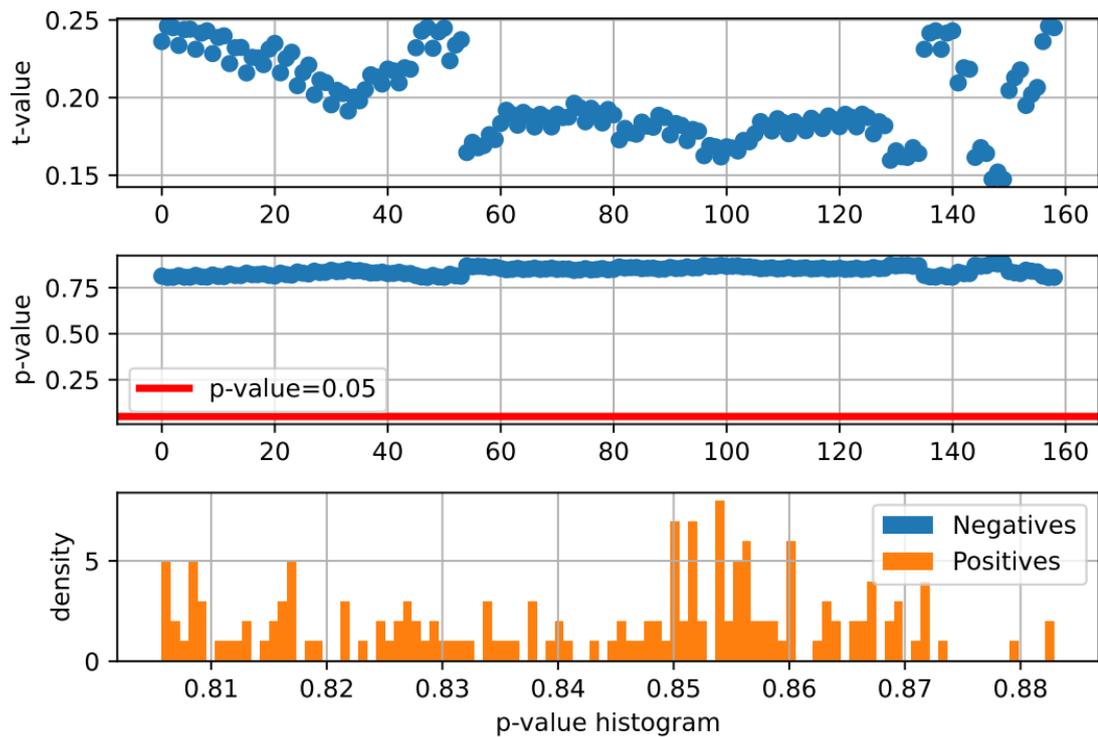
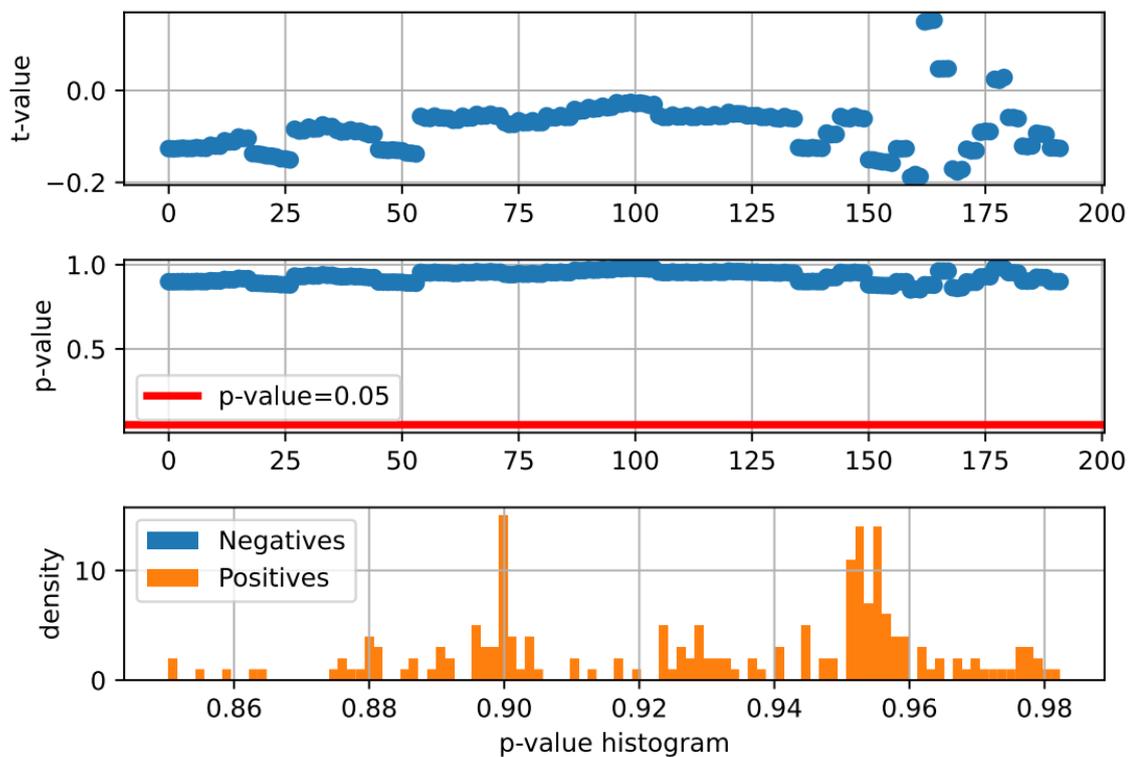


Figure 3.11: Results in terms of mean squared error of all Voltage phases  $\theta_{i,p}$  in islanded mode.

Figure 3.12: Test of the mean for all Voltage magnitudes  $V_{i,p}$ .Figure 3.13: Test of the mean for all Voltage phases  $\theta_{i,p}$

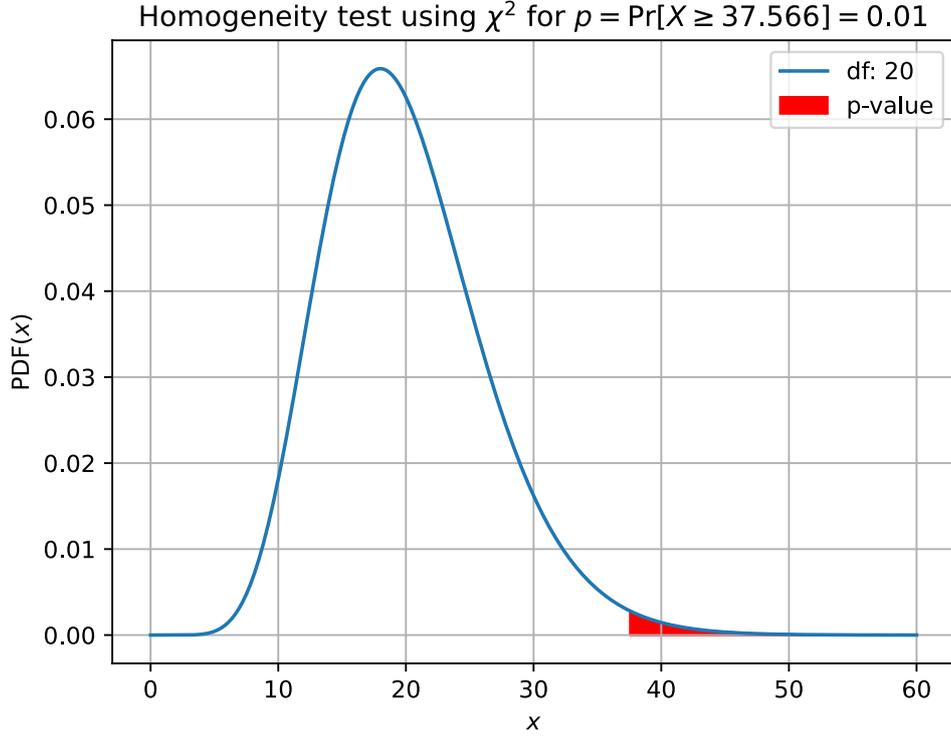


Figure 3.14: Homogeneity test for all state variables.

entry for  $0 \leq i < n$  is defined as follows:

$$\mathbf{x}[i] = \sum_{j=0}^m \mathbf{W}[i, j] \cdot \mathbf{z}[j] \quad (3.22)$$

Assuming that each operation can be done in  $O(1)$  time. The number of operations in the worst-case scenario is  $O(nm)$ :

$$O(nm) = \sum_{i=0}^n \sum_{j=0}^m 1 = nm \quad (3.23)$$

This implies that the worst-case computational complexity value of the proposed MLSE in running time is  $\Theta(nm)$ .

The advantages of the MLSE over the methods in the literature are: (i) Reduction of the implementation complexity, i.e., the formulation yields a unique constant matrix based on the trained weights to perform state estimation resulting in a non-iterative model. Hence, the number of iterations required to obtain a solution is one. This contrasts with the traditional WLS state estimation approach, which requires many iterations. Consequently, a significant computational time reduction is observed; (ii) versatility, the proposed approach can be used with any set of measurements. Furthermore, it is possible to run parallel implementations with models trained with different groups of measures in order to improve the reliability of the proposed approach; (iii) the MLSE can perform

state estimation in unobservable microgrids without pseudo-measurements, while Newton-like approaches do not; (iv) MLSE reports a stable behavior under heteroscedastic noise without explicitly knowing measurements' variances ( $\sigma$ ), while classical approaches require computing variances before performing state estimation.

## 3.7 Conclusions and Discussion

Machine learning state estimation reports high performance in heteroscedastic noisy measurements. The proposed machine learning state estimation is suitable for unobservable microgrids with low-cost smart meters. The MLSE model requires just a small number of samples to train a model. Only 500 power flow simulations were needed, in contrast with other approaches such as [66], [67] models that required 10 000 and 12 000 events for train a suitable model respectively, on the other hand the proposed approach, can handle heteroscedastic uncertainty on measurements. The proposed MLSE approach estimates the state without specify (i) an initial guess estimation point (a.k.a., *flat start*), nor (ii) an explicit definition of the measurement's variances ( $\sigma$ ), as is the case of the classic WLS approach. Herein, MLSE, is relevant for expanding existing knowledge in the state estimation area for three-phase, unbalanced and unobservable microgrid systems that are eligible for operating in grid-connected or islanding model. In this paper, gross errors are not considered at the available measurements for the state estimation. Future work will explore how to incorporate a pre-processing stage into the MLSE model to identify measures without gross errors, such as data latency and false data injection attacks, before performing state estimation.

The main difficulties of the proposed MLSE reside in the microgrid modeling used to perform precise simulations. A detailed model of each circuit component was necessary to accurately represent the microgrid's response to different scenarios. To overcome this challenge, the University of Campinas (UNICAMP) is currently deploying a project [1] to study and develop specialized models in the field of (i) power systems, (ii) renewable energy sources, (iii) power electronics, (iv) control systems, (v) optimization, and (vi) communications and information networks. A microgrid system modeled in OpenDSS with those specialized components was employed to perform this study.

# Chapter 4

## Contribution III

### Visual State Estimation for False Data Injection Detection of Solar Power Generation<sup>1</sup>

**Abstract:** As the penetration level of solar power generation increases in smart cities and microgrids, an automatic energy management system (EMS) without human supervision is most commonly deployed. Therefore, assuring safe and reliable data against cyber attacks such as false data injection attacks (FDIAs) has become of utmost importance. To address the aforementioned problem, this paper proposes detecting FDIAs considering visual data. The aim of visual state estimation is to enhance the resilience and security of renewable energy systems. This approach provides an additional layer of defense against cyber attacks, ensuring the integrity and reliability of solar power generation data and facilitating the efficient and secure operation of EMS. The proposed approach uses a modified VGG-16 neural network model to obtain an intermediate representation that provides textual and numerical explanations about the visual weather conditions from sky images. Numerical results and simulations corroborate the validity of our proposed approach. The performance of the modified VGG-16 neural network model is also compared with previous state-of-the-art machine learning models in terms of accuracy.

**Keywords:** solar power generation; false data injection attacks; computer vision; statistical approach.

---

<sup>1</sup>Contribution III is published as: **Byron Alejandro Acuña Acurio**, Diana Estefanía Chérrez Barragán, Juan Camilo López, Felipe Grijalva, Juan Carlos Rodríguez and Luiz Carlos Pereira da Silva, "Visual State Estimation for False Data Injection Detection of Solar Power Generation" [100] in Journal of Engineering Proceedings 2024, doi: <https://doi.org/10.3390/engproc2023047005>

## 4.1 Introduction

False data injection attacks (FDIAs), such as data poisoning or noise injection, can significantly affect the decision-making process of energy management system (EMS) applications, for instance, voltage regulation [101]. FDIAs are destructive to EMS [102], because the attacker can manipulate the meter readings by injecting additional false data, causing system instability [40] and even cascading failures leading to massive blackouts [103]. Therefore, various methodologies have been developed over the past decade to defend against such attacks. The existing methodologies can be divided into (i) protection-based and (ii) detection-based approaches [104]. The protection-based approaches [29, 105] are based on protecting specific sensors, but these methods have two drawbacks: (i) protecting data will reduce the amount of measured data, and (ii) the protection mechanism can not ensure that the data are always safe. On the other hand, FDIA detection methods traditionally were model-driven approaches. In [106], the authors proposed a method based on Kalman filters to detect FDIA in power grids. A multiobjective optimal detection scheme based on the parity method was proposed in [107], which only applies to DC microgrids but has poor adaptability to the current mainstream AC power grid. Although high detection accuracy is shown when tested on traditional power grid scenarios (without renewable energy resources), the literature does not answer whether the existing detection methods can be applied to microgrids with a high share of solar power generation plants. To the best of the authors' knowledge, the present work is the first attempt to apply computer vision techniques to the FDIA detection problem. However, for a detailed survey of FDIA detection strategies, the reader is referred to [108].

The fundamental problem in FDIA detection methods is to identify tampered measurement  $z_a$  reported from a smart metering system, which can be expressed as:

$$z_a = z + a. \quad (4.1)$$

In this case,  $a$  is a nonzero attack value added to the true measurement  $z$  of the solar power generation system. Facing this problem, this paper proposes an FDIA detection approach for solar power generation based on image processing of visual weather conditions using convolutional neural networks and transfer learning. To do this, a VGG-16 architecture pretrained on the ImageNet dataset was used. This work expands the theoretical explanation of transfer learning techniques to facilitate reproducibility by newcomers to this field. The main contributions of this work are:

- A novel deep learning architecture that can detect FDIA in solar power generation measurements based on sky images,

- A detailed step-by-step process to perform transfer learning from an object classifi-

cation domain to the FDIA detection domain in solar power generation.

The remainder of this paper is organized as follows. After this introduction, Section 4.2 explains each stage of the proposed methodology, step by step. In Section 4.3, the experimental setup is presented. Section 4.4 shows the numerical results. Finally, Section 4.5 presents the conclusions of this work.

## 4.2 Proposed Approach

This section describes the main components of our proposed method: (i) a transfer learning procedure using a modified VGG-16 convolutional neural network, (ii) an intermediate representation with a support vector regressor, and (iii) a binary hypothesis test.

### 4.2.1 Transfer Learning Procedure

The idea of transfer learning is to use a network previously trained with a large amount of data from a specific task and reuse it in a new task. In this case, a pretrained architecture named visual geometry group 16 (VGG-16) [2] was used as a starting point for our proposed FDAI approach. VGG-16 is a deep convolutional neural network (CNN) consisting of 16 layers with 1.2 million parameters. VGG-16 was initially pretrained using the ImageNet dataset that contains around 16 million images [2] to perform the classification task for 1000 different categories. In this work, the aforementioned VGG-16 architecture of the network was modified, discarding the fully connected and softmax blocks that are highlighted in red in Figure 4.1.

The convolutional and pooling layers shown in Figure 4.1 work as feature extractor layers; consequently, these layers have been frozen. The removed fully connected layers were replaced with the following layers in an ordered fashion: (i) a new batch normalization layer, (ii) a fully connected layer of 1024 neurons with a rectified linear unit (ReLU) as an activation function, (iii) a dropout layer with a rate set to 10% to avoid overfitting, and (iv) finally, a fully connected layer of 40 neurons with a sigmoid activation function. The original VGG-16 was trained to perform image classification (source domain  $\mathcal{D}_s$ ). A domain  $\mathcal{D}$  is the subject that performs learning. It consists of two parts: data  $(\mathcal{X}, \mathcal{Y})$  and the distribution  $P(\mathbf{x}, y)$  that generates such data for any sample  $(\mathbf{x}_i, y_i)$  in the available data  $\mathbf{x}_i \in \mathcal{X}, y_i \in \mathcal{Y}$ , as follows  $\mathcal{D} = \{\mathcal{X}, \mathcal{Y}, P(\mathbf{x}, y)\}$ .

In this case,  $\mathcal{D}_s$  was obtained using the ImageNet dataset that contains around 16 million images  $\mathcal{X}_s$  and 1000 different categories  $\mathcal{Y}_s$ . With transfer learning, it is possible to use the aforementioned modified VGG-16 for a new target domain  $\mathcal{D}_t = \{\mathbf{x}_j, y_j\}_{j=1}^{N_t}$ . Hence, in order to retrain the pretrained VGG-16 from the source domain to the target domain, the “*transient attribute* dataset”  $(\mathcal{X}_t, \mathcal{Y}_t)$  publicly available in [3] was used.

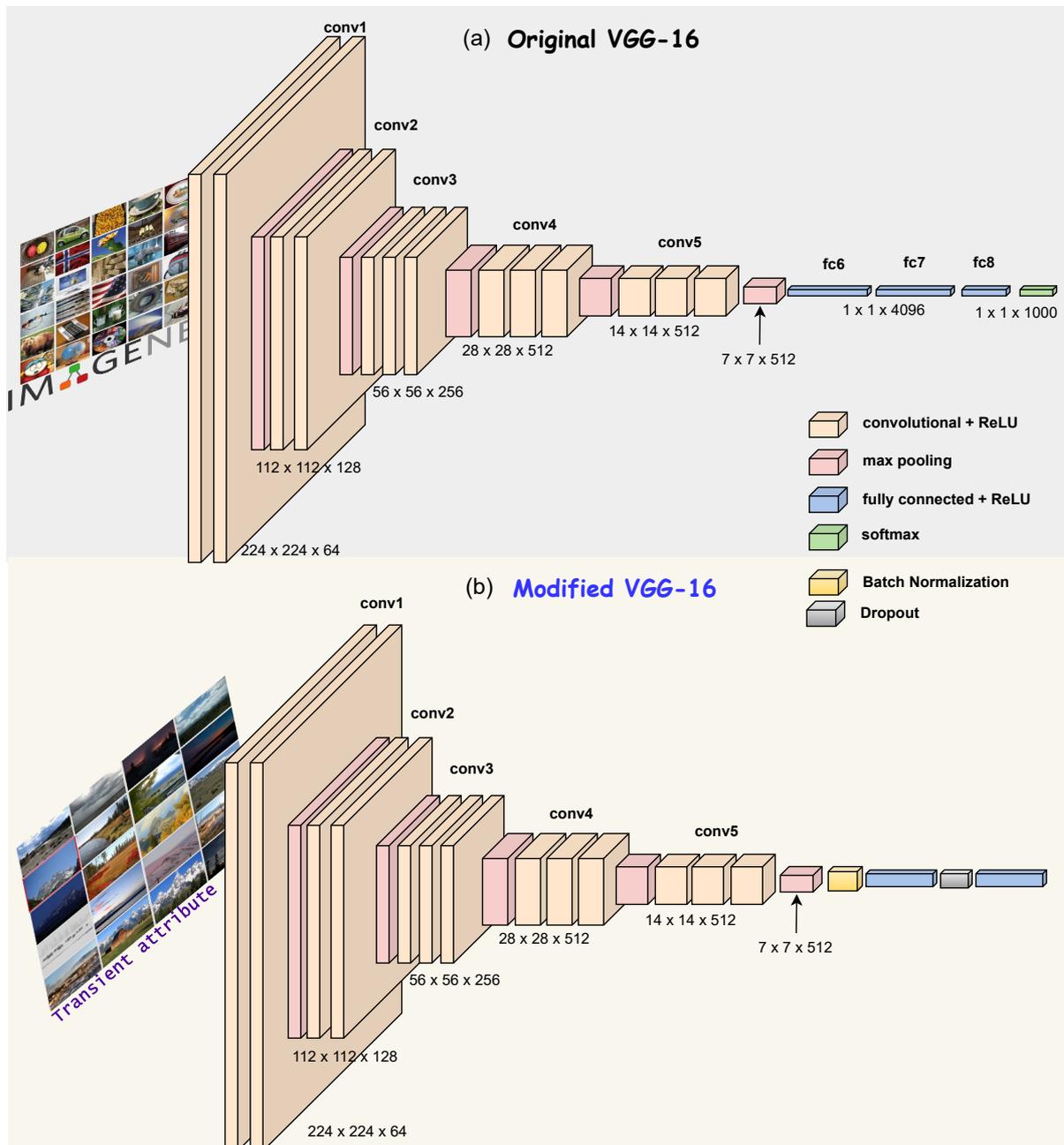


Figure 4.1: (a) Original VGG-16 [2]. (b) Modified VGG-16 employed to perform a transfer learning task.

The *transient attribute* dataset  $(\mathcal{X}_t, \mathcal{Y}_t)$ , has 8571 images  $\mathcal{X}_t$  from 101 webcams, all annotated with 40 attribute labels  $\mathcal{Y}_t$  of different types that contain values from 0 to 1. The 40 attributes are, for instance, lighting, the season of the year (winter, summer, etc.), weather (sunny, warm, cloudy, etc.), subjective impressions (beautiful, gloomy, soft, etc.), and some additional attributes such as dirty/polluted, busy, lush vegetation, etc. (see Figure 4.2 for some examples).

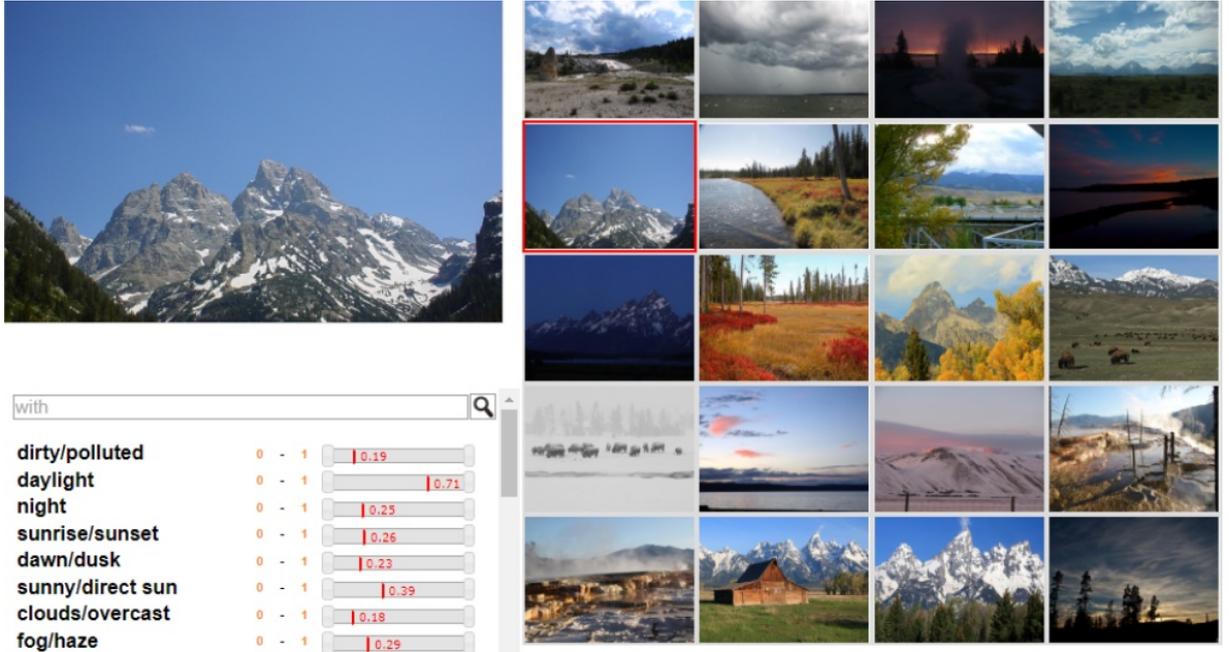


Figure 4.2: Example of an image from “transient attribute database  $\mathcal{X}_t, \mathcal{Y}_t$ ”, which is publicly available in [3] with their attribute label.

To perform the aforementioned transfer learning task, a finetuning of network weights was performed over all the layers of the modified VGG-16, since the images of the *transient attribute* dataset used for transfer learning are quite different from the ImageNet dataset used in the original VGG-16. In this work, the Adam optimization algorithm was used. Unlike the stochastic gradient descent (SGD), Adam can vary the learning rate throughout the training process to obtain a better performance model. The learning rate controls the variation of the network weights for each training epoch [109]. In this work, the initial learning rate was set at 0.001. If there was no network performance improvement during the training epochs, the learning rate was modified to 0.0005. This adjustment aimed to facilitate improvements until the final stages of training when there was no progress concerning its validation through the mean absolute error (MAE) over 15 consecutive epochs. Finally, the training process was limited to a maximum of 250 epochs. Given that the target domain  $\mathcal{D}_t$  is a 40 multi-output regressor, where each output ranges from 0 to 1, in this work, it was found that the most suitable loss function for performing the transfer learning task was the mean absolute error (MAE).

### 4.2.2 Intermediate Representation

Depending on the season of year and weather conditions, the modified VGG-16 neural network that was fitted using the aforementioned transfer learning approach looks for very different points in the image, as shown in the heat maps of Figure 4.3.

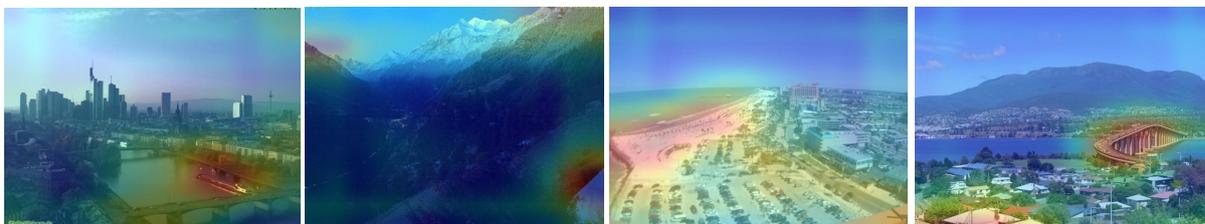


Figure 4.3: A heat map shows that the modified VGG-16 neural network looks for very different points in the image, depending on the weather conditions.

Therefore, the convolutional layers of the modified VGG-16 work as a feature extractor, as shown in Figure 4.4.

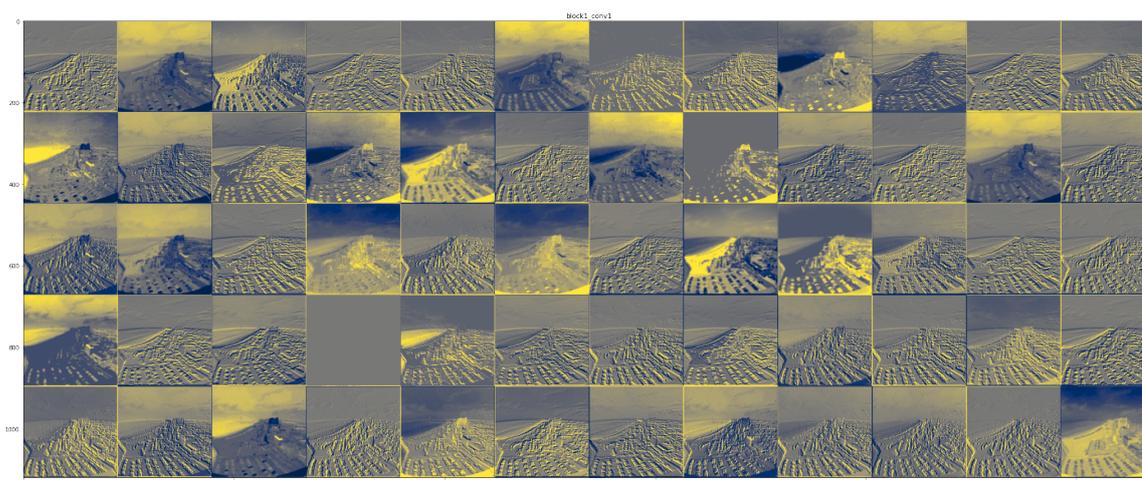


Figure 4.4: Example of features extracted by convolutional layers.

The output of the modified VGG-16 is 40 attributes, as shown in the correlation matrix in Figure 4.5.

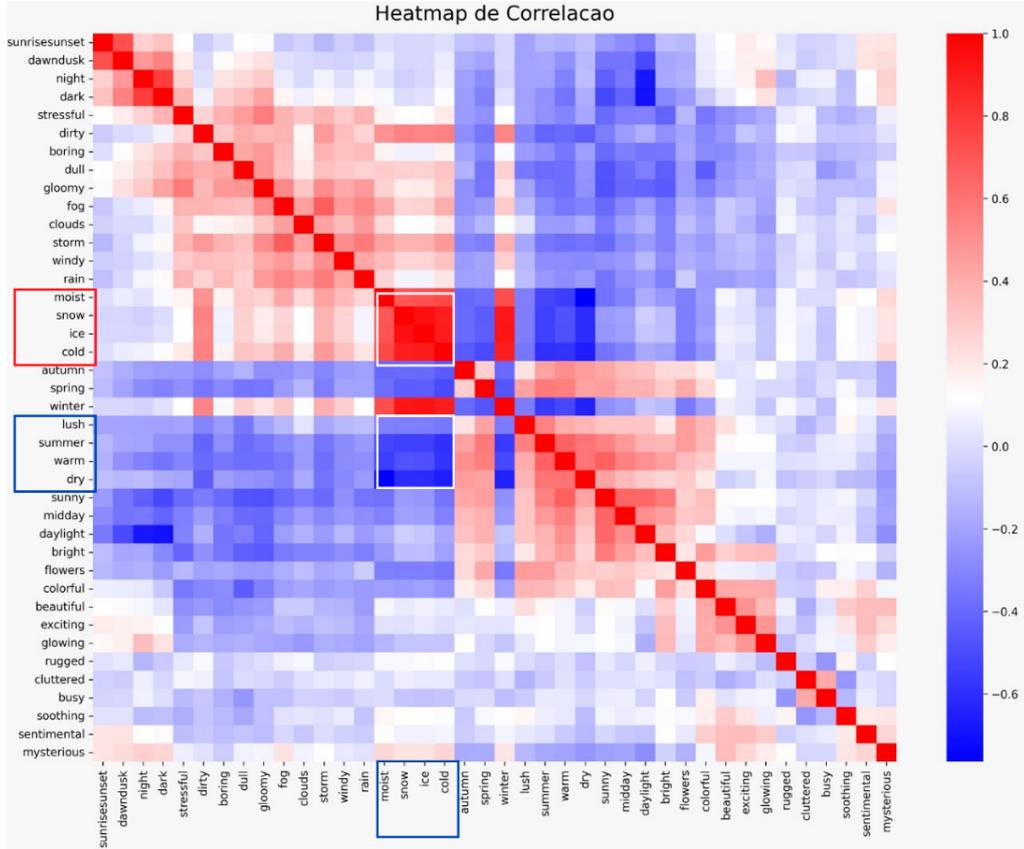


Figure 4.5: Correlation Matrix of the 40 attributes obtained from the modified VGG-16 from the images.

This work used these 40 attributes as an intermediate representation to perform the state estimation of solar power generation. The intermediate representation captures a numerical description of the variability in weather conditions that are very distinctive over time. To do this, historical sky images collected from the same location as the solar power generation plant under different conditions over one month were used to create a new dataset  $\mathcal{D}_{sky}$  of the historical intermediate representations  $\mathcal{X}_{sky}$  and the historical solar power generation  $\mathcal{Y}_{sky}$ . Then, a support vector regressor (SVR) with Radial Basis Function (RBF) kernel was trained using  $\mathcal{D}_{sky}$  to receive the intermediate representation and return the most likely state of solar power generation. It was empirically observed that 500 sky images were enough to fit the SVR.

### 4.2.3 Binary Hypothesis Test

To detect a false data injection attack on solar power generation measurements, an inequality chi-square test  $\chi^2$  was employed. For this, an estimated chi-square value  $\hat{\chi}^2$  was computed using the observed values and estimates, as follows:

$$\hat{\chi}^2 = \sum_{i=1}^m \frac{(O_i - E_i)^2}{E_i}, \quad (4.2)$$

where  $O$  represents the solar power generation measurements,  $E$  represents the estimated values of the solar power generation obtained from the SVR, and  $m$  is the number of measures over one day. To compute the chi-square probability distribution value  $\chi^2$ , the degrees of freedom  $k = m - n$  were defined, where  $n$  is the state variable, in this case  $n = 1$ . The estimated chi-square value  $\hat{\chi}^2$  was compared with the value of  $\chi^2$  for a given degree of freedom  $k$  and significance level  $\alpha$ . The significance level  $\alpha$  is the upper bound on the probability that a Type I error will occur after performing a hypothesis test. A Type I error occurs when the null hypothesis is correct but is rejected. In this work, a significance level of  $\alpha = 0.03$  was used, which indicates that there is a 3% chance that there are erroneous data or a confidence level of 97%. Consequently, the following chi-square test can be performed for the false data injection detection of solar power generation using visual state estimation:

If  $\hat{\chi}^2 \geq \chi_{k,\alpha}^2$ , false data injection attacks are suspected;

If  $\hat{\chi}^2 < \chi_{k,\alpha}^2$ , false data injection attacks are not suspected.

### 4.3 Experimental Setup

Images captured by the webcams have different sizes, requiring a preprocessing step before using the aforementioned approach. Initially, all images were resized to dimensions of  $200 \times 200$  with 3 color channels (RGB). Subsequently, for compatibility with the modified VGG16, the images were transformed into tensors of dimensions  $200 \times 200 \times 3$ .

The *transient attribute* dataset [3] with 8571 images was used to perform the transfer learning stage, to finetune the modified VGG-16. The *transient attribute* dataset was divided between testing and training data according to the original paper [3], but a double-stratified k-fold cross-validation approach was used to validate the modified VGG-16 during the training stage. Under this approach, the entire training dataset was divided into  $k = 10$  fold, one fold for validation and the remaining folds for training. This procedure allows us to obtain a more realistic idea of the performance of the model [109].

The modified VGG-16 neural network was trained with 250 epochs. However, empirically, it was observed that a good fit was obtained with only 25 epochs, as shown in the results section on the biases–variance curve, MAE metric, and R2 curve (see: Figure 4.6). Some data augmentation techniques were used in the finetuning stage of the modified VGG-16. The data augmentation stage aims to generate new images from the original ones. To achieve this goal, the original images were modified to generate new instances using the following set of transformations, which are:

*Rotation range:* This is the degree range for random rotations. In this study, a range between  $-5$  and  $+5$  degrees was used;

*Width shift range:* The original image is randomly shifted by a proportional percentage of the original image width. In this study, this parameter was set at 0.2, i.e., 20%;

*Height shift range:* This is similar to the previous transformation, but it uses the height of the image to perform the shift. In all the experiments, this parameter was set to 0.2, i.e., 20%;

*Zoom range:* This transformation generates a random zoom. In this work, a range between 90% and 110% of the original image was used;

*Horizontal flip:* This randomly flips the image horizontally;

*Vertical flip:* This randomly flips the image vertically;

*Brightness range:* This increases or decreases the brightness of the image. Hence, this parameter was set in a range between 0.9 and 1.1;

*Fill mode:* In all the experiments, the nearest approach fill mode was used, which fills points outside the boundaries of the image with similar information to that of the boundaries,

These transformations are applied to generate “on-demand” new images in the fine-tuning stage of the VGG-16.

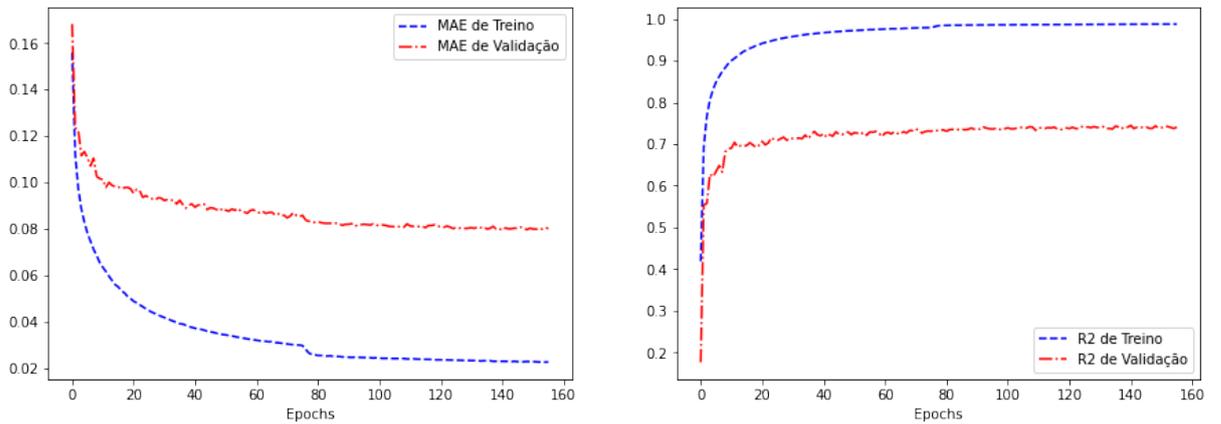


Figure 4.6: Numerical results of the training modified VGG-16 neural network.

## Svr Training Details

For reproducibility purposes, the publicly available dataset of sky images in [110] was used to create the new dataset  $\mathcal{D}_{sky}$  of historical intermediate representations  $\mathcal{X}_{sky}$  and historical solar power generation  $\mathcal{Y}_{sky}$ . This new dataset  $\mathcal{D}_{sky}$  was employed to train the support vector regression model.

## 4.4 Results And Discussion

Our proposed modified VGG-16 was compared to similar previous work [3]. To ensure a fair comparison the proposed model was trained using the same dataset that was reported in [3]. The results of our model are presented in Table 4.1. Figure 4.6 shows the training and validation learning curves of our proposed modified VGG-16 neural network model. Both the training and validation loss curves decreased after the modified VGG-16 model training began. This can be attributed to the transfer learning, indicating that the convolutional layer of the modified VGG-16 already had a high level of feature extraction, because these layers were pretrained. Although occasional fluctuations can be observed in the training loss, it is clear that the overall trend is a continuous decrease in the mean absolute error (MAE) during training.

Table 4.1: Comparison of the proposed modified VGG-16 neural network with previous works [3].

Modelo	MSE
SVM [3]	0.070
log reg [3]	0.093
SVR [3]	0.043
Proposed modified VGG-16	0.0319

The results of the intermediate representation show that the best-performing approach was accomplished using the proposed modified VGG-16 neural network as shown in Table 4.1.

The results indicate that, inside the critical region, the estimated values and the groundtruth values using  $\alpha = 0.03$ , as suggested in Section 4.2.3 for false data injection detection, can be considered the same, as shown in Figure 4.7.

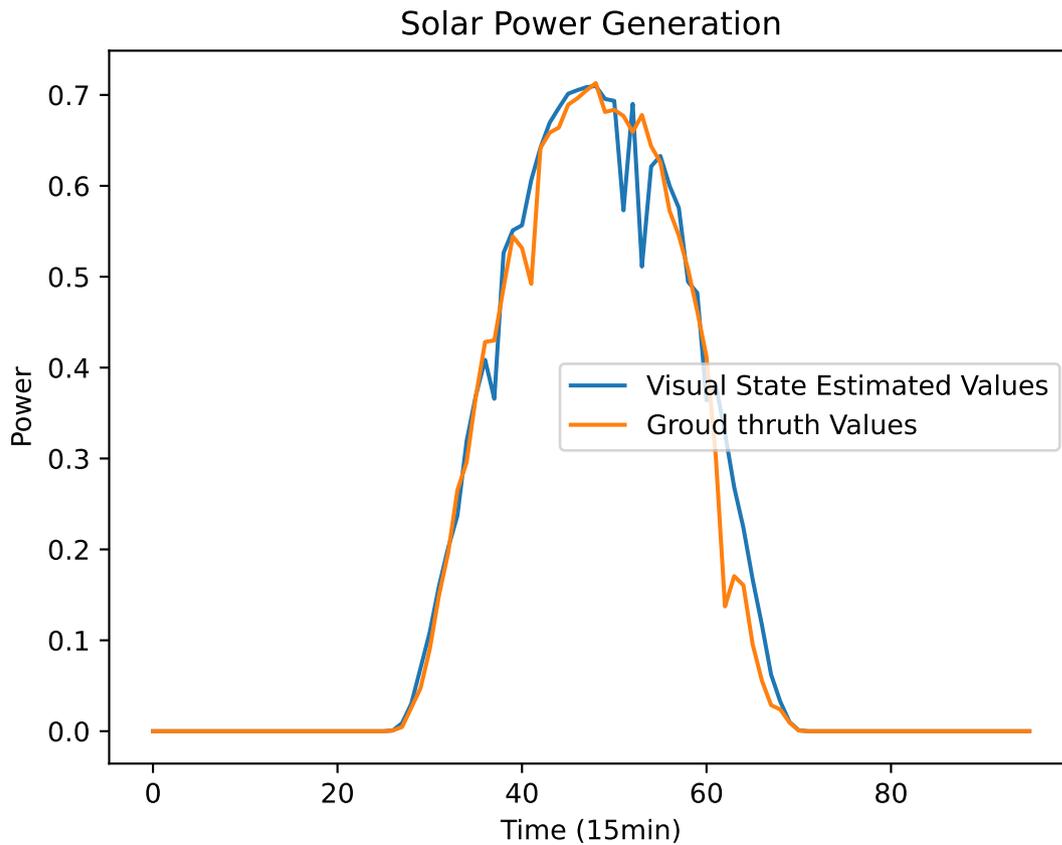


Figure 4.7: Comparison of the visual state estimated values and ground truth.

## 4.5 Conclusions

This work explores the false data injection detection in solar power generation from sky images, using a modified VGG-16 neural network to obtain an intermediate representation that can be used to estimate power generation with a support vector regressor. Results comparing the estimated values and the ground truth did not reveal a significant difference without a false data injection attack. A measurement discrepancy was detected when a data injection attack was performed in a random measurement. Our proposed approach overcomes the previous work [3] in terms of performance. The proposed approach is flexible and can be easily adapted to different solar power generation systems.

## 4.6 Future Works

As a future work, it could be interesting to study actions once a false data injection attack is detected; for instance, an appropriate mitigation strategy can be implemented. This might involve isolating the affected components, recalibrating sensors, restoring valid data from backups, or even triggering an automated response to neutralize the attack.

# Chapter 5

## Contribution IV

### Robust Data-Driven State of Health Estimation of Lithium-Ion Batteries Based on Reconstructed Signals <sup>1</sup>

**Abstract:** The state of health (SoH) of lithium-ion batteries is critical for diagnosing the actual capacity of the battery. Data-driven methods have achieved impressive accuracy, but their sensitivity to sensor noise, missing samples, and outliers remains a limitation for their deployment. This paper proposes a robust, purely data-driven SoH estimation methodology that addresses these challenges. Our method uses a proposed non-iterative closed-form signal reconstruction derived from a modified Tikhonov regularization. Five new features were extracted from reconstructed voltage and temperature discharge profiles. Finally, a Huber regression model is trained using these features for SoH estimation. Six ageing scenarios built from the public NASA and Sandia National Laboratories datasets, under severe Gaussian noise conditions (10 dB SNR), were employed to validate our proposed approach. In noisy environments and with limited training data, our proposed approach maintains a competitive accuracy across all scenarios, achieving low error metrics, with an RMSE on the order of  $10^{-4}$ , an MAE on the order of  $10^{-2}$ , and a MAPE below 1%. It outperforms state-of-the-art deep neural networks, direct-feature Huber models, and hybrid physics/data-driven models. In this work, we demonstrate that robustness in SoH estimation for lithium-ion batteries is influenced by the choice of machine learning architecture, loss function, feature selection, and signal reconstruction technique. In addition, we found that tracking the time to minimum discharge voltage and the time to maximum discharge temperature can be used as effective features to estimate SoH in data-driven models, as they are directly correlated with capacity loss and a decrease in power output.

---

<sup>1</sup>Contribution IV is published as: **Byron Alejandro Acuña Acurio**, Diana Estefanía Chérrez Barragán, Juan Carlos Rodríguez, Felipe Grijalva and Luiz Carlos Pereira da Silva, "Robust Data-Driven State of Health Estimation of Lithium-Ion Batteries Based on Reconstructed Signals" [111] in *Energies* 2025,

doi: <https://doi.org/10.3390/en18102459>

**Keywords:** Data-driven method; State of Health Estimation; Signal reconstruction; Regularization operator; Statistical features.

## 5.1 Introduction

Lithium-ion batteries (LIBs) are essential in the transition to sustainable energy systems [112]. They enable the storage of renewable energy for later use, which is critical for balancing supply and demand from intermittent energy sources [113]. In the electricity sector, LIBs are used for peak shaving, arbitrage, capacity firming, energy price management, frequency regulation, and grid stabilization [114]. Over time, batteries experience degradation due to multiple factors such as charge, discharge cycles, and the temperature effects of working conditions. State of health (SoH) estimation is important for assessing battery degradation of all the aforementioned applications and electric vehicles [115]. SoH is defined as the ratio of the actual capacity to the nominal capacity [116]. It is expressed as a percentage to diagnose the actual battery capacity [117]. For instance, a SoH below range of 70–80% indicates the end-of-life of the battery for primary applications since the battery can no longer deliver the required power density by high-performance storage systems [118]. However, such batteries can still be reused in second life applications (lower power applications) [119]. SoH estimation methods can be classified in the following four main categories [120]: (i) experimental methods, (ii) physics-based models, (iii) data-driven methods, and (iv) hybrid approaches [121].

Experimental methods, also known as direct measurement methods, include processes such as ampere-hour counting [122], which involves performing a full charge and discharge cycle between defined voltage thresholds. Specific current rates and temperature conditions are maintained during ampere-hour counting [120]. To determine the actual battery capacity  $Q$ , and SoH, the following formulation is employed:

$$\begin{aligned} Q &= \int_0^{t_f} \eta I(t) dt \\ SOH &= \frac{Q}{Q_{\text{initial}}} \end{aligned} \tag{5.1}$$

where  $Q$  is the actual capacity,  $Q_{\text{initial}}$  is the initial capacity of a battery,  $I(t)$  is the current, and  $t_f$  is the total time during charging or discharging.  $\eta$  represents the coulombic efficiency, which is very close to 1 for LIBs [123]. However, these methods are time-consuming, require interrupting normal operation, and are limited to controlled laboratory environments [124].

Physics-based models [125] simulate the chemical and physical processes occurring within batteries. SoH is estimated by monitoring changes in some model parameters that are correlated with battery aging [126]. For instance, in equivalent circuit models

(ECM) [127], LIBs are represented as electrical circuits with resistors, capacitors, and voltage sources [128]. As the battery ages, evidenced by a decline in SOH, the internal resistance parameter in ECM models increases, while the capacitance related to capacity tends to decrease [129]. Models calibrated under specific conditions may not apply directly to others, necessitating re-calibration across different operating scenarios [130].

Data-driven methods [131], such as machine learning (ML) algorithms, are trained on historical aging data (voltage, current, temperature, etc.) to identify complex patterns and relationships between the extracted features, e.g., health indicators and SoH. [132]. Recent studies highlight that by using large datasets and carefully designed training settings, data-driven techniques can achieve high accuracy and adaptability in predicting battery SoH [133], often outperforming traditional physics-based models in terms of accuracy [133]. Therefore, data-driven methods are usually implemented in BMS to track degradation and estimate SoH [134]. However, SoH data-driven approaches often require large datasets for training [135] and are sensitive to noise and missing data [136]. Thus, selecting and extracting the right features for data-driven methods is non-trivial, especially with varying operating conditions and degradation mechanisms [137].

Hybrid approaches combine different sources of information and modeling techniques to enhance the accuracy, robustness, and interpretability of SoH estimation. For instance, physics-based models can be used to provide simulated features to data-driven approaches to perform SoH estimation with real and simulated data [138]. Recent studies have proposed knowledge transfer techniques; for instance, [139] developed a transfer learning method that diagnoses degradation modes (DMs) of lithium–iron–phosphate (LFP) batteries by minimizing both classification loss and domain adaptation loss between synthetic and real datasets. This approach enables accurate DM identification without requiring extensive real-world labeled data. Similarly, [140] proposed a two-stage SOH estimation framework where DM knowledge is first transferred from synthetic to real datasets and then used as input for SOH prediction.

In this work, we propose a novel pure data-driven SoH estimation approach of lithium-ion batteries, designed to demonstrate robustness against noisy measurements and outliers. Unlike previous methods, our proposed approach can be trained with small datasets. The proposed method does not need an explicit physics-based model or assumptions about initial aging conditions for maintaining high accuracy, making it suitable for real-time BMS implementation with low-computational cost and memory requirements. To highlight the advantages of the proposed methodology, Table 5.1 presents a comparative analysis against previous data-driven and hybrid approaches that report the lowest error metrics for estimating lithium-ion battery state of health [141].

Table 5.1: Comparison of proposed SoH estimator with previous data-driven approaches.

Data-Driven Method	Number of Input Features	Robustness	Trainable Parameters	Performance Metric
Proposed Approach	5 proposed features based on signal reconstruction	Our proposed signal reconstruction approach can handle noise and outliers in the measurement data	6 polynomial parameters trained with Huber cost function	RMSE = $10^{-4}\%$ , MAE = $10^{-2}\%$ , MAPE = 1%
Deep Neural Network [5]	3 (Direct features)	The reported approach requires preparing the data by removing significant outliers manually	2 hidden layers with 30 and 15 neurons, respectively, as well as Sigmoid and Tanh activation functions	RMSE = $1.9 \times 10^{-4}\%$ , MAPE = 1.39%
Deep Neural Network [142]	6 (Direct features)	The paper does not discuss a dedicated noise handling mechanism	217 trainable parameters	RMSE = 0.004758%, MAE = 0.534%
Nonlinear Autoregressive Exogenous Neural Network [143]	8 (Model-based features)	The paper does not discuss a dedicated noise handling mechanism	Hidden neurons = 50, Feedback delays = 8	MAE = 0.72%, MaxE = 4.69%
Gated Recurrent Unit Network [144]	3 (Direct features)	Gaussian noise injection with a mean of 0 and a standard deviation of 1–2% into the voltage, current, and temperature measurements (works on less noise-corrupted signals)	Hidden neurons = 256 (GRU), convolution number = 64, size of each convolution layer $32 \times 1$	MAE = 1.03%, MaxE = 4.11%
Convolutional Neural Network [145]	1 (Preprocessed features)	The reported approach is sensitive to noise and outliers	Number of convolution kernels = 256, size of the kernel = $3 \times 1$	RMSE = 1.1%, MAE = 0.9%

The summarized contributions of this paper are as follows:

Closed-form signal reconstruction: We present a non-iterative closed-form solution for the signal reconstruction of noisy measurements.

Novel data-driven health indicators: We introduce five noise-resilient features derived from the reconstructed voltage and temperature discharge profiles.

Robust data-driven state of health estimation: In this work, we use the Huber cost function to improve the accuracy of the regression model by reducing the impact of outliers, providing an alternative to removing outliers from the dataset.

This paper is organized as follows: Section 5.1 introduces the SoH estimation problem for LIBs and reviews current modeling techniques in the literature. Section 5.2 outlines the proposed approach and the case studies. Section 5.3 presents a comprehensive performance assessment that demonstrates the effectiveness of our method compared to existing data-driven and hybrid methods in noisy environments and with limited training

data. Finally, Section 5.4 concludes the paper, summarizing key findings and discussing potential directions for future work.

## 5.2 Materials and Methods

This section outlines the materials and methodology adopted in the development of the proposed SoH estimation approach for LIBs under noisy measurement conditions. As illustrated in Figure 5.1, the proposed method is described in detail in Section 5.2.1. This work assumed that there is an availability of data storage systems capable of continuously monitoring and recording voltage and temperature measurements from LIBs.

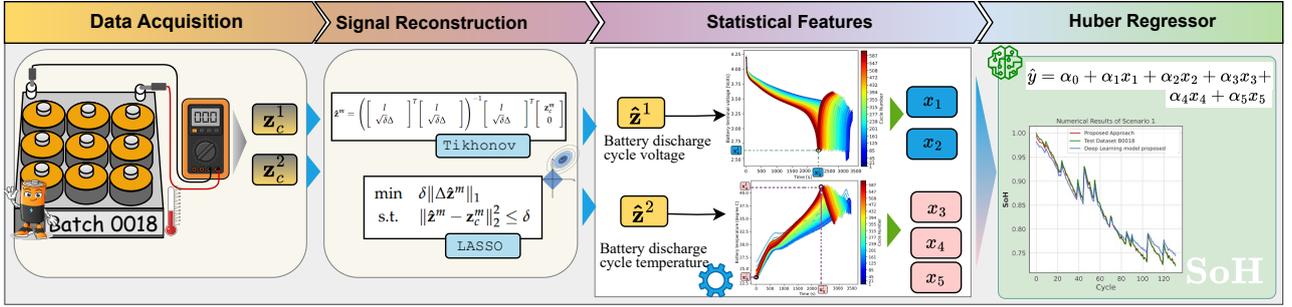


Figure 5.1: Proposed state of health estimation for lithium-ion batteries based on signal reconstruction.

### 5.2.1 Proposed Approach

Our proposed approach consists of three key stages:

1. **Signal Reconstruction:** This stage estimates a vector  $\hat{\mathbf{z}}^m = (\hat{z}_1^m, \hat{z}_2^m, \dots, \hat{z}_t^m)^T$  from noisy voltage and temperature profiles  $\mathbf{z}_c^m = (z_{c,1}^m, z_{c,2}^m, \dots, z_{c,t}^m)^T$ , where  $m = 1$  corresponds to the battery discharge voltage profile and  $m = 2$  represents the battery discharge temperature profile. The signal reconstruction employs the proposed closed-form, non-iterative mathematical expression formulated in Equation (5.8).
2. **Feature Extraction:** From the reconstructed voltage and temperature discharge profiles, five new data-driven health indicators are extracted, which are conditionally correlated with the battery aging process.
3. **SoH Estimation:** A Huber regression model was employed for SoH estimation, demonstrating robustness against outliers.

Further details of each stage will be explained in the following sections.

### 5.2.1.1 Proposed Signal Reconstruction stage

This subsection details the proposed signal reconstruction stage, designed to recover voltage and temperature discharge profiles from noisy measurement data. The measurement model employed in this work was formulated as follows:

$$\mathbf{z}_c^m = \mathbf{z}^m + \mathbf{e} \quad (5.2)$$

where  $\mathbf{z}^m$  represents the unknown noise-free signal for each discharge profile. The term  $\mathbf{e}$  denotes the stochastic noise component affecting the measurements. The objective of the proposed signal reconstruction method is to estimate the clean signal, denoted as  $\hat{\mathbf{z}}^m$ , from the noisy measurement  $\mathbf{z}_c^m$ . To achieve this, the signal reconstruction problem is formulated as a regularized optimization problem:

$$\min_{\hat{\mathbf{z}}^m} \|\hat{\mathbf{z}}^m - \mathbf{z}_c^m\|_2^2 + \delta \phi(\Delta \hat{\mathbf{z}}^m); \quad (5.3)$$

the regularization hyperparameter  $\delta$  controls the trade-off between preserving the original noisy signal and noise filtering. The function  $\phi(\Delta \hat{\mathbf{z}}^m)$  introduces a regularization term to improve robustness against measurement noise. When  $\delta = 0$ , the reconstructed signal exactly matches the noisy measurement. Conversely, excessively high values of  $\delta$  result in over-smoothing, potentially distorting the original signal characteristics. Two regularization strategies were investigated in this study: (i) Tikhonov regularization =  $\phi(\Delta \hat{\mathbf{z}}^m) = \|\Delta \hat{\mathbf{z}}^m\|_2^2$  and (ii) LASSO regularization =  $\phi(\Delta \hat{\mathbf{z}}^m) = \|\Delta \hat{\mathbf{z}}^m\|_1$ .

To solve the ill-posed problem in Equation (5.3), the regularization hyperparameter  $\delta$  is determined using the L-Curve method, as detailed in [40]. Additionally, we employ the regularization operator  $\Delta$ , previously reported in [40], and defined in Equation (5.4), due to its demonstrated stability in reconstructing voltage and temperature discharge profiles of lithium-ion batteries observed in this work.

$$\Delta = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & -1 & 1 \end{bmatrix} \quad (5.4)$$

where  $\Delta \in \mathbb{R}^{t \times t}$  is a square matrix, with its dimension equal to the number of samples  $t$  in each discharge profile.

**Signal Reconstruction Based on Tikhonov Regularization** The reconstruction of the signal is formulated based on a modified Tikhonov regularization strategy expressed in Equation (5.5). This formulation, known as the unrestricted form, allows us to obtain a non-iterative approach to recover clean discharge profiles from noisy measurements, as follows

$$\min_{\hat{\mathbf{z}}^m} \|\hat{\mathbf{z}}^m - \mathbf{z}_c^m\|_2^2 + \delta \|\Delta\hat{\mathbf{z}}^m\|_2^2 \quad (5.5)$$

This formulation can be rewritten in the following compact form:

$$\min_{\hat{\mathbf{z}}^m} \|\hat{\mathbf{z}}^m - \mathbf{z}_c^m\|_2^2 + \delta \|\Delta\hat{\mathbf{z}}^m\|_2^2 = \min_{\hat{\mathbf{z}}^m} \left\| \begin{bmatrix} \hat{\mathbf{z}}^m - \mathbf{z}_c^m \\ \sqrt{\delta} \Delta\hat{\mathbf{z}}^m \end{bmatrix} \right\|_2^2 \quad (5.6)$$

Equation (5.6) can be rewritten as

$$\min_{\hat{\mathbf{z}}^m} \left\| \begin{bmatrix} \hat{\mathbf{z}}^m - \mathbf{z}_c^m \\ \sqrt{\delta} \Delta\hat{\mathbf{z}}^m \end{bmatrix} \right\|_2^2 = \min_{\hat{\mathbf{z}}^m} \left\| \begin{bmatrix} I \\ \sqrt{\delta} \Delta \end{bmatrix} \hat{\mathbf{z}}^m - \begin{bmatrix} \mathbf{z}_c^m \\ \mathbf{0} \end{bmatrix} \right\|_2^2 \quad (5.7)$$

Solving for  $\hat{\mathbf{z}}^m$ , the closed-form solution of signal reconstruction using modified Tikhonov regularization is given by

$$\hat{\mathbf{z}}^m = \left( \begin{bmatrix} I \\ \sqrt{\delta} \Delta \end{bmatrix}^T \begin{bmatrix} I \\ \sqrt{\delta} \Delta \end{bmatrix} \right)^{-1} \begin{bmatrix} I \\ \sqrt{\delta} \Delta \end{bmatrix}^T \begin{bmatrix} \mathbf{z}_c^m \\ 0 \end{bmatrix} \quad (5.8)$$

**Signal Reconstruction Based on LASSO Regularization** In the case of LASSO regularization strategy, it is not possible to have a closed-form expression. However, the ill-posed problem can be effectively solved using the following constrained optimization problem.

$$\begin{aligned} \min \quad & \delta \|\Delta\hat{\mathbf{z}}^m\|_1 \\ \text{s.t.} \quad & \|\hat{\mathbf{z}}^m - \mathbf{z}_c^m\|_2^2 \leq \delta \end{aligned} \quad (5.9)$$

For this study, the regularization parameter  $\delta$  was empirically set to 5 for signals with a signal-to-noise ratio (SNR) of 10 dB. The optimal choice of  $\delta$  is inherently dependent on the noise characteristics of the data:

For highly corrupted signals (SNR < 10 dB), larger  $\delta$  values ( $\delta > 5$ ) are recommended to enhance noise suppression.

For signals with minimal noise contamination (SNR > 10 dB), lower  $\delta$  values ( $\delta < 5$ ) preserve original signal details.

### 5.2.1.2 Proposed Feature Extraction stage

We propose a systematic approach to extract features highly correlated with battery health degradation by analyzing the evolving patterns in voltage and temperature profiles during discharge cycles. The method consists of three steps:

1. Identification of discharge cycle.
2. Extraction of voltage features.
3. Extraction of temperature features.

1. Identification of discharge cycle: We isolate the continuous discharge cycles from battery operation data by identifying periods where the current drops below a predefined threshold (in this case  $-0.05$  A) and where a predefined minimum voltage (e.g., cutoff voltage) is reached. Note that manufacturers generally set a safe cutoff voltage to preserve battery health and lifespan [146]. Lower cutoff voltages (e.g., 2.0 V) correspond to a depth of discharge (DOD) closer to 100%. However, in real-time applications, battery manufacturers generally do not recommend a 100% DOD as an operational norm [147]. Many battery studies suggest limiting the DOD to 70% to extend battery lifespan [148]. This step ensures we analyze only discharge events, filtering out partial or interrupted discharge cycles that could skew the analysis.

2. Extraction of voltage features: Battery voltage during discharge follows a decreasing pattern, where each successive voltage measurement is lower than the previous one,  $\hat{z}_1^1 > \hat{z}_2^1 > \dots > \hat{z}_t^1$ , and whose characteristics change as the battery degrades. We capture these dynamics through two critical features:

Minimum discharge voltage ( $x_1$ ): This is the lowest voltage reached during the discharge cycle (cutoff voltage). As a battery ages, its internal impedance increases due to factors such as lithium inventory loss and conductive degradation [149]. An increment in the impedance leads to a more pronounced voltage drop ( $R_0 i$ ) during discharge [150]. According to the Shepherd model, the terminal voltage  $V$  at discharge time  $t$  can be modeled as follows [151]:

$$V(t) = E_0 - K \left[ \frac{Q}{Q - i(t) \times t} \right] i(t) - R_0 i(t) \quad (5.10)$$

where  $E_0$  is the theoretical initial open-circuit voltage under specific working conditions,  $R_0$  represents the internal ohmic resistance of the battery ( $\Omega$ ),  $i$  is the discharge current in amperes (A), assumed positive, and  $Q$  is the actual full capacity of the battery in ampere-hours (Ahr), in which eventually, with aging, its

experiment capacity will fade.  $Q - i(t) \times t$  is the remaining capacity in the battery at discharge time  $t$  and  $K$  is the polarization resistance coefficient ( $\Omega$ ). Thus, the minimum discharge voltage is sensitive to internal resistance growth, and to the battery's full capacity  $Q$ . In this work, we found that tracking the time to minimum discharge voltage can be used as a feature to estimate SoH in data-driven models.

Time to minimum voltage ( $x_2$ ): This is the duration between the beginning of discharge and the minimum voltage. A healthy battery with full capacity can sustain the discharge current longer before reaching the voltage limit, whereas an aged battery with capacity fade will hit the minimum voltage sooner [152].

3. Extraction of temperature features: Temperature profiles during discharge typically exhibit an increasing pattern, where each successive temperature reading surpasses the previous one,  $\hat{z}_1^2 < \hat{z}_2^2 < \dots < \hat{z}_t^2$ , and which changes with battery aging. To capture these thermal characteristics, we introduce three temperature-related features:

Minimum temperature at the beginning of discharge ( $x_3$ ): The baseline temperature at the beginning of discharge, establishing a reference point.

Maximum discharge temperature ( $x_4$ ): The peak temperature reached during discharge, reflecting internal resistance and exothermic reactions. As the battery degrades and its internal resistance increases, it produces more heat for the same discharge current, resulting in a higher peak temperature [153].

Time elapsed between minimum and maximum temperature ( $x_5$ ): The aged battery's temperature climbs to its maximum in a shorter time than in a new battery, which heats more slowly due to its lower internal resistance [154].

These proposed features effectively encapsulate the electrochemical and thermal signatures of battery degradation without requiring complete charge–discharge curves. All extracted features  $\mathbf{x}_i$  were normalized to lie between 0 and 1 to ensure numerical stability and consistency.

### 5.2.1.3 State of Health (SoH) Estimation

SoH is estimated using a Huber regression model (Equation (5.11)), a robust method less sensitive to outliers than least squares regression [155]. The Huber loss function [155] (Equation (5.12)) is used to fit the model, which smoothly transitions from a quadratic form for small residuals to a linear form for large residuals. This adaptive property mitigates the influence of outliers while maintaining sensitivity to minor deviations [155], making it particularly well-suited for SoH estimation of noisy data.

$$\hat{y} = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 + \alpha_5 x_5 \quad (5.11)$$

where the  $\alpha$  coefficients are optimized by solving a convex quadratic programming problem [156] and  $x_i$  represents our proposed features. We set the Huber transition parameter to  $\gamma = 1.35$ , which yields approximately 95% efficiency under Gaussian noise. This choice promotes numerical stability by limiting the influence of extreme residuals, thereby helping to prevent overfitting even in high-dimensional feature spaces. The Huber cost function [157], used in the training stage to fit the model, is defined as follows:

$$L_\gamma(y_i, \hat{y}_i) = \begin{cases} \frac{1}{2}(y_i - \hat{y}_i)^2 & \text{for } |y_i - \hat{y}_i| \leq \gamma, \\ \gamma |y_i - \hat{y}_i| - \frac{1}{2}\gamma^2 & \text{otherwise.} \end{cases} \quad (5.12)$$

Once trained, the Huber regression model provides accurate SoH estimates,  $\hat{y}$ , from the five extracted features. As shown in Figure 5.1, incorporating Huber regression into our pipeline ensures computational efficiency and robustness against outliers.

Training and testing were performed using labeled datasets, which are described comprehensively in Section 5.2.2.

## 5.2.2 Datasets

The proposed SoH estimation method was evaluated using two distinct datasets: the NASA battery dataset [7] and Sandia National Laboratories (SNL) battery degradation dataset [8].

### 5.2.2.1 NASA Data Overview

To evaluate our proposed SoH estimation method, experiments were performed using three distinct testing scenarios derived from NASA’s lithium-ion battery aging dataset [7]. The selected batteries and scenarios ensure various realistic operational scenarios, including temperature variations, different discharge cutoff voltages, diverse capacity fade thresholds, and the end-of-life (EOL) criteria.

Scenario 1: Variable Temperature Conditions. In this scenario, we evaluate our proposed SoH estimation method over a broad temperature range from 4 °C to 40 °C. Batteries B0005 and B0007 were used for model training, while Battery B0018 was reserved exclusively for testing purposes. The charging protocol consisted of a constant current (CC) phase at 1.5 A until the battery voltage reached 4.2 V, then a constant voltage (CV) phase was maintained at 4.2 V until the charging current decreased to 20 mA. The discharge cutoff voltages varied across the batteries, set at 2.7 V for B0005, 2.2 V for B0007, and 2.5 V for B0018. The end-of-life criterion was determined based on a capacity fade of 30%, corresponding to a decrease from the nominal capacity of 2.0 Ah to 1.4 Ah.

Scenario 2: Constant Temperature Conditions. In Scenario 2, the batteries operated

under controlled ambient conditions, with a constant temperature of 24 °C. Battery B0033 was used for training purposes, while B0034 was reserved for testing. The charging protocol was identical to Scenario 1, with the difference that the discharge cutoff voltages were set at 2.0 V for B0033 and 2.2 V for B0034. A more demanding end-of-life criterion was adopted, defined by a capacity fade threshold of 20%, also corresponding to a reduction in battery capacity from the initial 2.0 Ah to 1.6 Ah.

**Scenario 3: Low-Temperature Operation:** In this scenario, the batteries operated under low-temperature conditions (4 °C). The charging protocol followed the previously established procedure, with the difference that the discharge cutoff voltages among batteries were set at the following: 2.2 V for B0046, 2.5 V for B0047, and 2.7 V for B0048. The end-of-life criterion was the same as Scenario 1, defined as a 30% capacity fade.

Using the aforementioned scenarios, we evaluate the proposed SoH estimation method under different working conditions, as described in [158] and shown in Table 5.2.

Table 5.2: Operating parameters of batteries from NASA dataset.

Battery ID	End Voltage (V)	Ambient Temperature (°C)	Nominal Capacity (Ah)	Discharge Current (A)	End of Life Criteria (Ah)	No. of Cycles
B0005	2.7	4 to 40	2	2	1.4	168
B0007	2.2	4 to 40	2	2	1.4	168
B0018	2.5	4 to 40	2	2	1.4	132
B0033	2.0	24	2	4	1.6	198
B0034	2.2	24	2	4	1.6	198
B0046	2.2	4	2	1	1.4	72
B0047	2.5	4	2	1	1.4	72
B0048	2.7	4	2	1	1.4	72

### 5.2.2.2 SNL Data Overview

We also evaluated our proposed SoH estimation method using three additional testing scenarios derived from the Sandia National Laboratories (SNL) lithium-ion battery aging dataset [8]. SNL’s dataset contains detailed cycle-level and time-series battery performance data collected during charge–discharge cycling tests on commercial 18650-format lithium–iron–phosphate (LFP) cells. Each battery was subjected to varied experimental conditions, including temperature, charge–discharge current rates, and state-of-charge (SoC) ranges. The data were acquired through the open-access web platform [159].

**Scenario 4: High-stress regime.** In this scenario, both training and test datasets from lithium–iron–phosphate (LFP) 18650 cells were cycled at 25 °C ambient temperature. Cells were charged at a rate of 0.5C and discharged at a rate of 1C (where C represents the battery’s capacity in ampere-hours (Ah)), covering a full SoC range from 0% to 100%

(complete depth of discharge). Specifically, the training data originated from the file `SNL_18650_LFP_25C_0-100_0.5-1C_c_timeseries.csv`, while testing data were taken from `SNL_18650_LFP_25C_0-100_0.5-1C_d_timeseries.csv`. Discharge rate conditions varied between 0.5C and 2C.

Scenario 5: Higher discharge rate. The training dataset was obtained from the file `SNL_18650_LFP_15C_0-100_0.5-2C_a_timeseries.csv`, while the testing dataset was sourced from `SNL_18650_LFP_15C_0-100_0.5-2C_b_timeseries.csv`. Compared to other scenarios, Scenario 5 is characterized by its lower operating temperature conditions (15 °C) and a higher discharge rate (2C), resulting in shorter discharge durations at high voltage per cycle.

Scenario 6: High-stress regime. The working conditions of this scenario are similar to those of Scenario 4, but with different batteries. For the training data, we used the file `SNL_18650_LFP_25C_0-100_0.5-1C_a_timeseries.csv`, while the data used for testing were taken from the file `SNL_18650_LFP_25C_0-100_0.5-1C_b_timeseries.csv`.

The proposed SoH estimation methodology was evaluated across the aforementioned defined scenarios, detailed in Table 5.3. Note, NCA means Lithium Nickel Cobalt Aluminum Oxide, and NMC means Lithium Nickel Manganese Cobalt Oxide. Hence, NCA and NMC refer to two different types of lithium-ion battery cathode chemistries.

Table 5.3: Operating parameters of batteries from SNL dataset.

	Cathode Chemistry	
	NCA	NMC
Manufacturer	Panasonic	LG Chem
Manufacturer City - Country	Osaka - Japan	Seoul - South Korea
Manufacturer PN	NCR18650B	18650HG2
Battery type	18650	18650
Nominal capacity [Ah]	3.2	3
Nominal voltage [V]	3.6	3.6
Voltage range [V]	2.5–4.2	2.0–4.2
Max discharge current [A]	6	20
Temperature range [°C]	0–45	–5–50
Charge C-rate	0.5C	0.5C
Discharge C-rate	0.5C/1C/2C	0.5C/1C/2C
Test temperature [°C]	15/25/35	15/25/35
Depth of discharge	0–100%	0–100%

### 5.2.3 Accuracy Evaluation Metrics

A robust SoH estimation model maintains its accuracy across a wide range of scenarios. To evaluate the accuracy of our proposed SoH estimation methodology, three widely used evaluation metrics [160] were employed: mean absolute percentage error (MAPE), mean absolute error (MAE), and root mean square error (RMSE). These metrics provide

complementary measures of estimation accuracy, as detailed below:

Mean Absolute Percentage Error (MAPE) quantifies the relative error as a percentage, making it easy to interpret and compare across different datasets or models, and is defined as follows:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (5.13)$$

Note that MAPE can be misleading when actual  $y_i$  values are close to zero, as the percentage error becomes disproportionately large.

Mean Absolute Error (MAE) measures the average magnitude of absolute deviations between the estimated and actual SoH values. Lower MAE values indicate higher prediction accuracy. Mathematically, it is expressed as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (5.14)$$

Note that MAE is less sensitive to outliers compared to metrics like Root Mean Squared Error (RMSE), as it does not square the error.

Root Mean Square Error (RMSE) evaluates the standard deviation of prediction errors, heavily penalizing large deviations. RMSE is computed as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (5.15)$$

Note that RMSE is sensitive to outliers.

where  $\hat{y}_i$  is the estimated SoH,  $y_i$  is the ground truth SoH, and  $n$  is the total number of SoH observations.

## 5.3 Numerical Results and Discussions

We first describe the experimental setup in Section 5.3.1, then provide a comprehensive experimental validation and performance analysis of our SoH estimation methodology under various noise levels, charging protocols, and operational conditions.

### 5.3.1 Experimental Setup

Our proposed SoH estimation method was evaluated using the scenarios detailed in Section 5.2.2. All experiments were conducted on a workstation assembled in Dell's manufacturing facility located in Hortolândia, São Paulo, Brazil, with an Intel(R) Core(TM) i7 870 processor and 8 GB RAM. The proposed approach presented in Section 5.2.1 was implemented in Python 3.8 [161].

### 5.3.2 Sensitivity Analysis of Signal Reconstruction

The voltage and temperature discharge profile data for each scenario, detailed in Section 5.2.2, were distorted according to the following signal-to-noise ratio (SNR) ranges: 10, 20, 30, 40, 50 dB. All signals were normalized to zero mean and unit variance before processing to ensure a fair comparison across methods and parameter values. For each SNR level, we identified the optimal regularization parameter  $\delta$  for both Lasso and Tikhonov methods using grid search. Our proposed approach assumes that sensor noise follows a zero-mean Gaussian distribution. This is a common characteristic of real-world sensor data according to [162]. Both our proposed approach based on modified Tikhonov formulation (Equation (5.8)) and LASSO formulation (Equation (5.9)) demonstrated stable performance under noisy conditions. However, our proposed approach achieves the lowest error metrics across multiple SNR levels, outperforming filtering methods such as moving average (MA) [163] and Kalman filter [164] for signal reconstruction, as shown in Figure 5.2.

In contrast to our proposed closed-form solution, which provides stable reconstructions for ill-posed inverse problems, LASSO formulation requires an iterative solution process and can be sensitive to convergence criteria and regularization parameter tuning. Therefore, we found that the LASSO formulation has a higher error variability than our proposed approach based on modified Tikhonov. On the other hand, the moving average (MA) filter operates by smoothing data points over a specified window. However, MA can lead to oversmoothing and the loss of important features in the signal, particularly in the presence of sharp transitions or edges. We found that this characteristic often results in insufficient performance when the signal reconstruction involves non-stationary or rapidly changing signals. In contrast, our proposed approach promotes piecewise constant solutions that are beneficial for reconstructing signals with abrupt changes. Thus, our proposed signal reconstruction method not only reduces noise effectively but also preserves essential signal characteristics that simple averaging methods can overlook. In this study, the centered MA filter [163] was implemented using Equation (5.16)

$$y[n] = \frac{1}{M} \sum_{i=0}^{M-1} x \left[ n + i - \frac{M-1}{2} \right] \quad (5.16)$$

where  $M$  represents the predefined samples (window size) that are used to average the original input signal  $x[n]$  and  $n$  is the time index. In this work, we use  $M = 5$  obtained from a grid search. This allows the filter to process both past and future samples in a symmetric manner.

Kalman filter addresses the signal reconstruction problem by forming a recursive esti-

mation procedure that integrates predictions based on the following dynamic model [164]:

$$x_k = x_{k-1} + w_k, \quad w_k \sim \mathcal{N}(0, Q), \quad (5.17)$$

where  $x_k \in \mathbb{R}$  is the latent (noise-free) signal sample at discrete time step  $k$  and  $Q > 0$  is the process-noise variance that captures unmodeled perturbations and model mismatch. Each noisy observation  $z_k$  was modeled with direct noisy measurement  $v_k$ , as follows:

$$z_k = x_k + v_k, \quad v_k \sim \mathcal{N}(0, R), \quad (5.18)$$

where  $R > 0$  is the measurement-noise variance. Our proposed approach based on Tikhonov formulation uses all data (e.g., all discharge voltage and temperature profiles) in a single batch. In contrast, the Kalman filter performs a sequential filtering by updating the reconstruction solution at each sample  $k$ . The Kalman filter achieves optimality only when  $(Q, R)$  match the true process and sensor variances. Hence, we found that our proposed method is more robust to non-optimal hyperparameter specification compared with the Kalman filter. The experimental results demonstrate that our proposed approach achieves lower bias and variance than the Kalman filter for the signal reconstruction task under varying noise levels. This highlights the importance of sufficient data information for accurate reconstruction, providing stability against noise. Figure 5.2 is organized as follows: the boxplots show the RMSE, MAE, and MAPE for all the estimated SoH across different SNRs. SNR quantifies the measurement error for each measurement using (5.19).

$$\text{SNR}_{[\text{dB}]} = 10 \log_{10} \frac{[z^2]}{\sigma_z^2} \quad (5.19)$$

The error  $\mathbf{e} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma_z^2)$  is used to pollute each measurement  $\mathbf{z}_c^m$  in Equation (5.2). Signal reconstruction is an ill-conditioned problem, which can lead to convergence issues. To mitigate this, we introduced a closed-form solution in Equation (5.8) that ensures reliable performance even when the LASSO formulation encounters convergence difficulties. To do this, it is required to use the proposed regularization operator presented in Equation (5.4), which was introduced in a previous paper by our research group [40]. Since the proposed closed-form expression is a non-iterative mathematical model, the number of iterations required to obtain a signal reconstruction solution is one. Consequently, a significant reduction in computational time is observed.

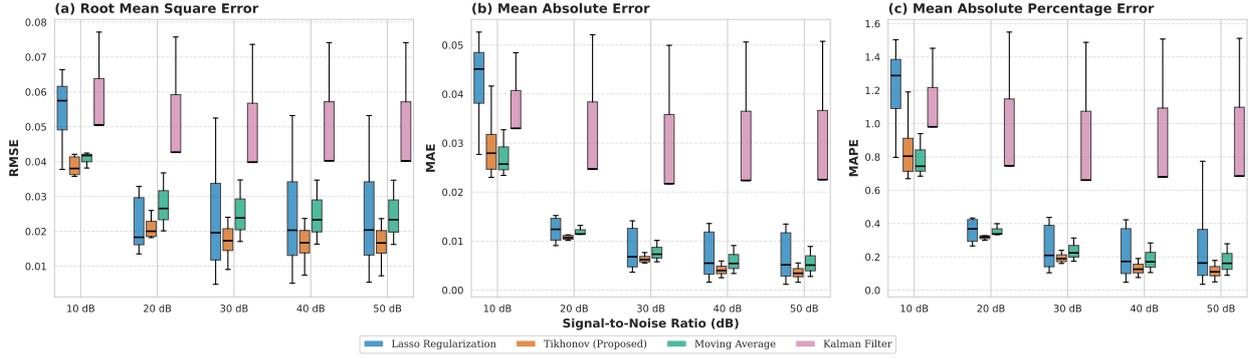


Figure 5.2: Sensitivity analysis of signal reconstruction quality across various SNRs (10–50 dB). Our proposed approach based on Tikhonov consistently outperforms widely adopted signal processing methods, including the LASSO formulation, moving average (MA) filter, and the Kalman filter, across multiple accuracy metrics. Boxplots show (a) RMSE, (b) MAE, and (c) MAPE, with the proposed approach yielding the lowest error dispersion across all metrics, highlighting its effectiveness in preserving signal quality under noise.

Table 5.4 shows the reduction in reconstruction signal quality as noise levels increase. However, our proposed approach maintains a consistent good performance in terms of accuracy. These findings support the selection of our proposed modified Tikhonov reconstruction stage within the proposed SoH estimation method.

Table 5.4: Comparison of signal reconstruction accuracy for the proposed closed-form Tikhonov, the iterative LASSO formulation, a centred moving-average (MA) filter, and a recursive Kalman filter under five noise conditions (SNR = 10–50 dB). Lower values of RMSE, MAE, and MAPE indicate superior reconstruction quality. The best results in each column are in bold.

Metrics	Method	SNR (dB)					Average
		10	20	30	40	50	
RMSE	Tikhonov	<b>0.0404</b>	<b>0.0210</b>	<b>0.0172</b>	<b>0.0165</b>	<b>0.0164</b>	<b>0.0223</b>
	MA	0.0408	0.0278	0.0252	0.0248	0.0248	0.0287
	Lasso	0.0547	0.0248	0.0240	0.0247	0.0248	0.0306
	Kalman	0.0594	0.0537	0.0511	0.0515	0.0515	0.0534
MAE	Tikhonov	0.0295	<b>0.0112</b>	<b>0.0064</b>	<b>0.0041</b>	<b>0.0035</b>	<b>0.0109</b>
	MA	<b>0.0273</b>	0.0120	0.0077	0.0060	0.0056	0.0117
	Lasso	0.0427	0.0144	0.0104	0.0091	0.0088	0.0171
	Kalman	0.0381	0.0338	0.0311	0.0318	0.0319	0.0334
MAPE (%)	Tikhonov	0.8497	<b>0.3295</b>	<b>0.1940</b>	<b>0.1299</b>	<b>0.1133</b>	<b>0.3233</b>
	MA	<b>0.7894</b>	0.3566	0.2369	0.1866	0.1764	0.3492
	Lasso	1.2197	0.4241	0.3164	0.2809	0.2719	0.5026
	Kalman	1.1377	1.0139	0.9360	0.9553	0.9601	1.0006

### 5.3.3 Sensitivity and Correlation Analysis of Proposed Features

It is necessary to understand the importance of each feature for SoH estimation and provide insights into model goodness-of-fit in relative terms; that is, not merely whether the model performs well, but which features contribute most to its performance. To do this, we conducted a sensitivity analysis of the proposed features shown in Table 5.5. In this analysis, we examined model performance under different feature combinations: using all features, leaving one feature out, and using only a single feature. We evaluated each combination using error metrics (RMSE, MAE, MAPE), the coefficient of determination ( $R^2$ ), and execution time. Note that  $R^2$  quantifies the proportion of variance in the SoH variable that is estimated from one or more features, defined as

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5.20)$$

where  $\bar{y}$  is the mean of the true values. In this formulation,  $R^2 = 1$  indicates a perfect fit of the model to the data, corresponding to zero residual error ( $\sum_{i=1}^n (y_i - \hat{y}_i)^2 = 0$ ), whereas  $R^2 = 0$  indicates that the model is no better than predicting the mean of the data ( $\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \bar{y})^2$ ). Thus,  $R^2 = 0.90$  means that 90% of the variance in SoH is explained by the model. It is worth noting that  $R^2$  can be negative if the model's predictions are worse than simply using the mean  $\bar{y}$ , although in a well-trained regression for SoH we expect  $R^2$  to be between 0 and 1.

We use the Pearson correlation coefficient to measure the degree of linear correlation between our proposed features and the SoH, based on the following formulation

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (5.21)$$

where  $x_i$  indicates the proposed features and  $y_i$  the SoH, and  $\bar{x}_i$  and  $\bar{y}_i$  are their average values, respectively. The range of the Pearson correlation coefficient is  $[-1, 1]$ . The closer to the extreme values at both ends, the stronger the linear correlation between the proposed features and SoH. For detailed information of the correlation analysis, see Table 5.6.

Table 5.5: Sensitivity analysis of proposed features. Checkmark ( ✓ ) indicates inclusion and a dash (-) indicates exclusion of the corresponding proposed features: minimum discharge voltage ( $x_1$ ), time to minimum voltage ( $x_2$ ), minimum temperature at the start of discharge ( $x_3$ ), maximum discharge temperature ( $x_4$ ), and time elapsed between minimum and maximum temperature ( $x_5$ ). The best results are in bold.

Proposed Features					RMSE	MAE	MAPE (%)	$R^2$	Time (s)
$x_1$	$x_2$	$x_3$	$x_4$	$x_5$					
✓	✓	✓	✓	✓	<b>0.0862</b>	<b>0.0732</b>	<b>0.5987</b>	0.923523	0.0259
-	✓	✓	✓	✓	0.233401	0.1855705	1.2169	0.913716	0.01905
✓	-	✓	✓	✓	1.0651395	0.8038405	2.60585	0.628143	0.0176
✓	✓	-	✓	✓	0.2290405	0.181089	0.70005	0.922905	0.01785
✓	✓	✓	-	✓	0.230986	0.18282	0.64605	0.9241995	0.01835
✓	✓	✓	✓	-	0.2297205	0.181146	0.73265	0.923476	0.01975
✓	-	-	-	-	1.393372	1.171515	8.9394	-1.454	0.0076
-	✓	-	-	-	0.231793	0.1860235	1.22655	<b>0.9275</b>	0.0086
-	-	✓	-	-	1.1681775	0.862342	4.66665	0.08354	0.0079
-	-	-	✓	-	1.1566025	0.8757905	6.4991	-0.63832	<b>0.0067</b>
-	-	-	-	✓	1.364249	1.157678	3.33345	0.4079215	0.00895

Table 5.6: Pearson correlation coefficients ( $r$ ) between each of the five proposed features and the battery state of health (SoH), reported separately for the training set, the independent test set, and the combined dataset. The strong positive correlations of the temporal features  $x_2$  and  $x_5$  with SoH corroborate their dominant explanatory power, whereas  $x_1$ ,  $x_3$ , and  $x_4$  show only weak or moderate association. A negative sign indicates an inverse monotonic relationship with SoH.

Proposed Features	Train	Test	Overall
$x_1$ : minimum discharge voltage	-0.1465	-0.2116	-0.1791
$x_2$ : time to minimum voltage	0.9689	0.9627	0.9658
$x_3$ : minimum temperature at the start of discharge	-0.0340	0.2752	0.1206
$x_4$ : maximum discharge temperature	0.0357	-0.0601	-0.0122
$x_5$ : time elapsed between minimum and maximum temperature	0.9625	0.9305	0.9465

Based on Table 5.6, we found that tracking the time to minimum discharge voltage and the time to maximum discharge temperature can be used as effective features to estimate SoH in data-driven models, as they are directly correlated with capacity loss and a decrease in power output.

### 5.3.4 Comparison of Feature Selection: Direct Features vs Proposed

We compared our proposed five features, detailed in Section 5.2.1.2, with the direct features approach reported in [4], which uses the following ten features: minimum, maximum, and average values of voltage, current, temperature, and the total discharge time. Both methods use Huber regression, which is robust to outliers, along with the proposed signal reconstruction stage, differing only in their features. As depicted in Figure 5.3, our proposed method outperformed the direct features method in terms of accuracy across different battery-aging scenarios. Our proposed features, despite using fewer features than direct features [4], perform better, showing that our proposed feature engineering is more effective than just collecting a wide range of statistics.

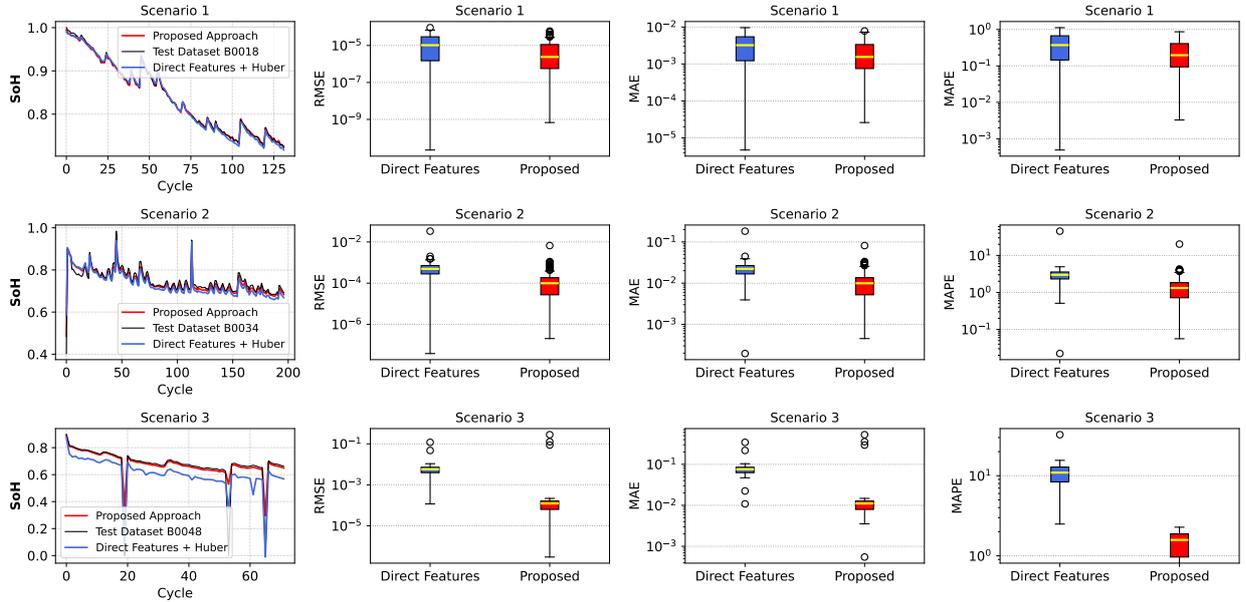


Figure 5.3: Performance comparison of state-of-health (SoH) estimation across three scenarios using direct features reported in [4] vs. our proposed feature selection approach. In each scenario, the left plot shows predicted SoH at different numbers of aging cycles. The boxplots show RMSE, MAE, and MAPE. The proposed method consistently exhibits lower prediction errors, highlighting its effectiveness over direct features.

### 5.3.5 Data-Driven Robustness Comparison

To evaluate the robustness [162] of our proposed approach, we adjusted the experimental setup presented in [165] with more intense noise contamination. The original study [165] used Gaussian white noise with an SNR of 30 dB. We evaluated our model with an SNR = 10 dB. This simulated more adverse data conditions. We compared our proposed approach with the deep neural network (DNN) model for SoH estimation. The DNN model is described in [5] and summarized in Section 5.3.5.1. To eliminate any bias

that might arise from differences in noise handling, we employed the same proposed signal reconstruction stage based on modified Tikhonov regularization for the DNN model, which was used in our proposed SoH estimation approach. Note that SoH estimations were performed across different battery aging stages, operational conditions, and charging protocols.

#### 5.3.5.1 DNN

The DNN model extracts three key temporal features from the constant current–constant voltage (CC-CV) charging process: (1) the initial voltage inflection point, which characterizes the early charging behavior; (2) the CC-CV transition time, occurring when the cell reaches 4.2 V and the current decreases to 1.5 A; and (3) the time to reach peak cell temperature, which captures thermal characteristics during charging. The DNN architecture consists of two hidden layers with three neurons each, employing Rectified Linear Unit (ReLU) activation functions between layers.

As illustrated in Figure 5.4, our proposed approach consistently estimates the SoH throughout the battery life cycles. The temporary increment in the measured value of SoH (non-linearities) in Figure 5.4 corresponds to the capacity regeneration phenomenon that occurs in lithium-ion batteries [166]. Our approach demonstrates superior robustness against noise, particularly in Scenario 2, where the presence of high noise levels significantly affects the DNN model [5]. While DNN models require large datasets for effective learning, our method achieves comparable or superior performance with significantly fewer training samples in noisy conditions. In addition, in Scenario 3, which includes multiple missing values in the training data, our method maintained its resilience, further demonstrating its robustness under adverse conditions.

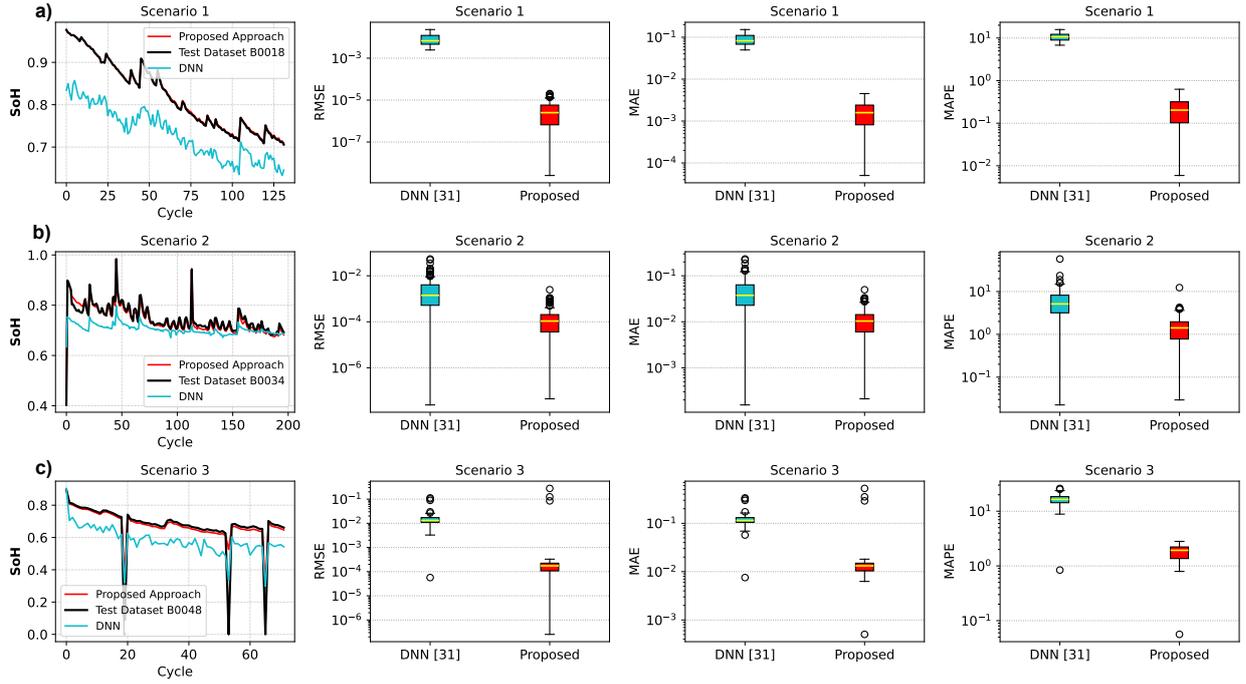


Figure 5.4: (a) In Scenario 1, our proposed approach consistently estimates the SoH throughout the battery life cycle. (b) In Scenario 2, the comparison highlights the robustness of our proposed approach against noisy data, whereas the DNN approach [5] is sensitive to noise. (c) In Scenario 3, the results demonstrate the resilience of our proposed approach to some missing values.

Compared to the DNN [5] model, our proposed approach achieves superior performance (see Figure 5.4) with a RMSE of  $10^{-4}$ , MAE of  $10^{-2}$ , and MAPE below of 1%. These results confirm the ability of our proposed approach to perform consistent and accurate SoH estimations across different discharge conditions and with limited training data. We further benchmarked our results against prior state-of-the-art methods, as shown in Table 5.1.

### 5.3.6 Comparison with Hybrid Model

Our proposed approach is a purely data-driven machine learning (ML) model. We compare our proposed approach with the hybrid approach in [6], which combines ML-based predictions with a physics-based model. In the test of the hybrid approach, we employed the proposed pipeline in Section 5.2.1 as the machine learning component. For the physics-based model, an exponential decay assumption was employed, as follows:

$$\text{SoH}_{\text{phys}}(n) = C_f + (C_0 - C_f) \exp(-\beta n^\alpha) \quad (5.22)$$

where  $C_0$  represents the initial capacity (normalized to 1.0),  $C_f$  is the final capacity (e.g., 0.8), and  $\beta, \alpha$  are decay parameters that control how quickly the SoH degrades with cycle count. As shown in Figures 5.5 and 5.6, the hybrid approach model might

not perfectly capture real-world battery dynamics. This limitation underscores that our proposed data-driven method can help correct or improve these SoH estimations. To compute the following results, hybrid model fusion was weighted 30% physics and 70% ML as follows:

$$\text{SoH}_{\text{hybrid}} = w \cdot \text{SoH}_{\text{phys}} + (1 - w) \cdot \text{SoH}_{\text{ML}} \quad (5.23)$$

The hybrid approach was tested using the six scenarios presented in Section 5.2.2.

Table 5.7 compares the accuracy of the proposed SoH estimation pipeline with three alternatives over six representative aging scenarios. These results confirm that our proposed pipeline estimates SoH, even under severe noise, missing data, and variable discharge protocols, where the alternative methods lose accuracy.

As is shown in Table 5.8, SoH estimation of lithium-ion batteries is influenced by operational parameters such as discharge C-rates, temperature, and charging protocols. In this case, higher discharge C-rates tend to larger estimation errors. However, our proposed approach maintains competitive accuracy. Low temperatures like 4 °C lead to higher errors compared to moderate temperatures such as 24 °C. The CC-CV charging protocol performs better than CC for SoH estimation.

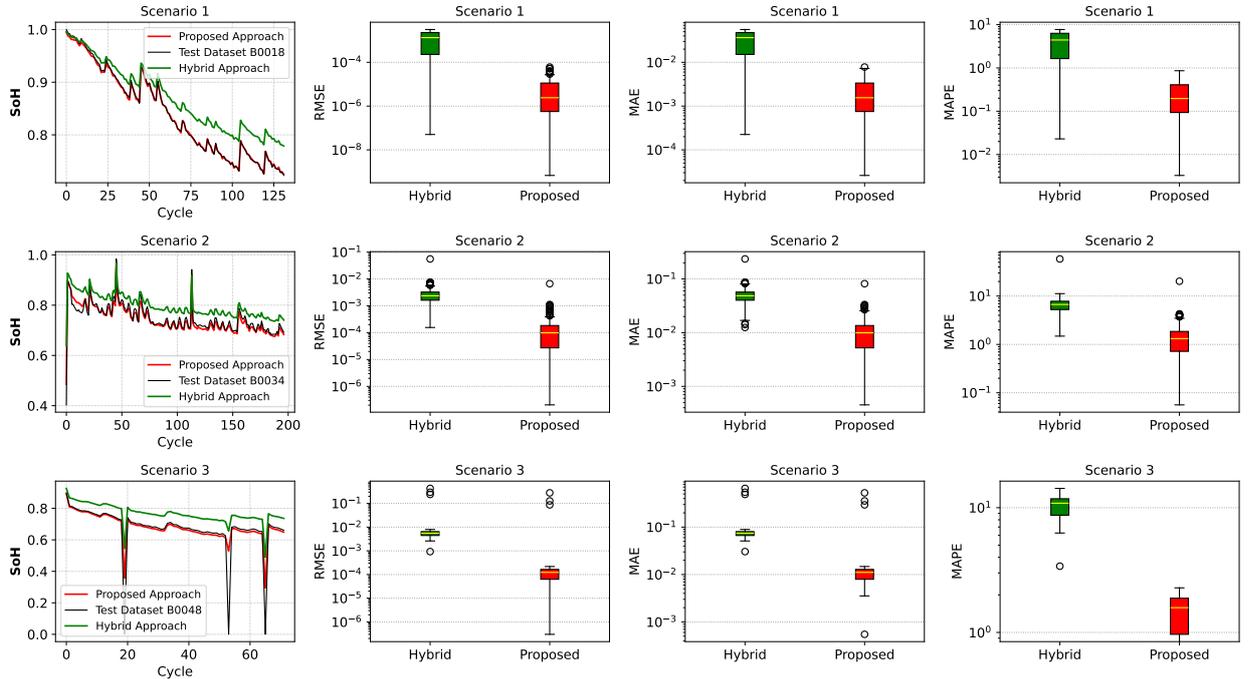


Figure 5.5: Comparison of proposed SoH approach (red) versus hybrid model (green) [6] across multiple scenarios based on NASA dataset [7] and detailed in Section 5.2.2. Left plots show the SoH predictions for each scenario along different numbers of aging cycles (the ground truth is in black). Boxplots of the corresponding error metrics RMSE, MAE, and MAPE. Overall, our proposed purely data-driven approach captures real-world battery aging dynamics more accurately and consistently than the hybrid model.

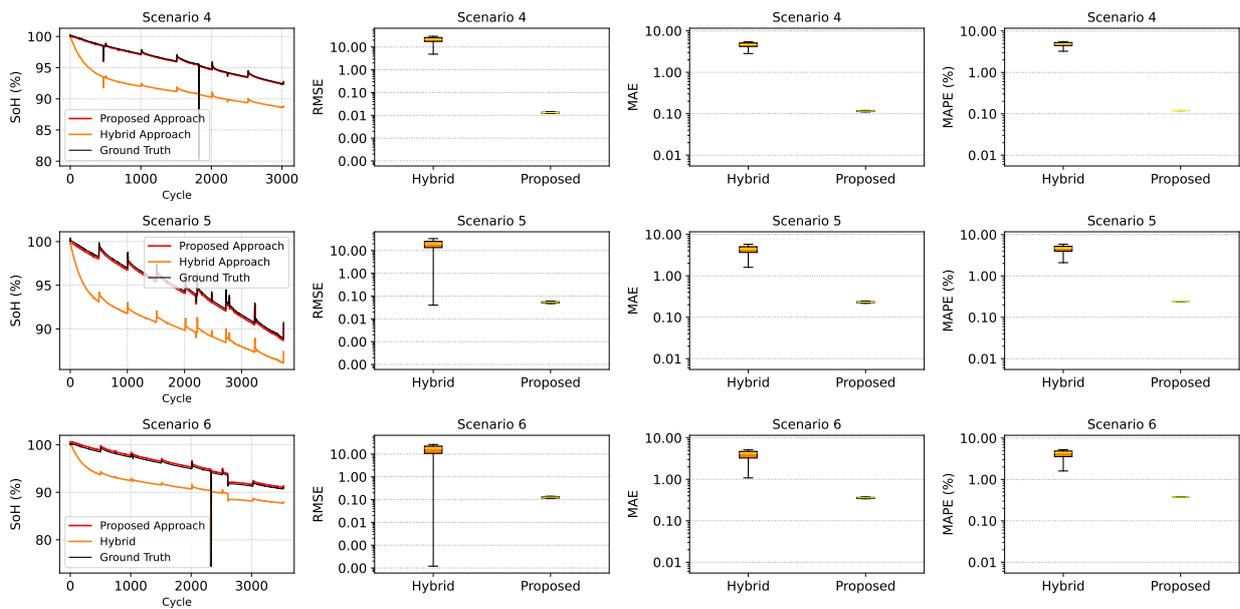


Figure 5.6: Comparison of proposed SoH approach (red) versus hybrid model (orange) [6], across multiple scenarios based on SNL dataset [8] and detailed in Section 5.2.2. Left plots show the SoH predictions for each scenario along different numbers of aging cycles (the ground truth is in black). Boxplots of the corresponding error metrics RMSE, MAE, and MAPE. We can see our proposed pure data-driven approach can adapt to the nuances of battery aging in ways the hybrid approach could not.

Table 5.7: Comparison of proposed state-of-health (SoH) estimator with a deep neural network (DNN) using three handcrafted charging features, a Huber-regression model trained on ten direct statistics, and a hybrid estimator (physics-aided). Performance is evaluated under six aging scenarios, with results reported as root mean square error (RMS), mean absolute error (MAE), and mean absolute percentage error (MAPE). Lower values indicate higher accuracy. Aggregated results are labeled “Average”. The best results are in bold.

Scenario	Method	RMSE	MAE	MAPE (%)
1	Proposed	<b>0.0029</b>	<b>0.0022</b>	<b>0.2546</b>
	DNN	0.018239	0.015416	1.9341
	Direct	0.0042	0.0034	0.4231
	Features			
	Hybrid	2.277851	1.922806	2.5100
2	Proposed	<b>0.0140</b>	<b>0.0110</b>	<b>1.4985</b>
	DNN	0.103690	0.097540	12.8129
	Direct	0.0274	0.0238	3.2878
	Features			
	Hybrid	2.845887	1.890512	2.6018
3	Proposed	<b>0.0835</b>	<b>0.0261</b>	<b>1.4191</b>
	DNN	0.178465	0.174136	23.8429
	Direct	0.0883	0.0793	10.6426
	Features			
	Hybrid	10.205681	3.207709	1.7226
4	Proposed	<b>0.0640</b>	<b>0.0567</b>	<b>0.06</b>
	DNN	9.3996	9.3972	9.78
	Direct	4.5677	0.5558	0.58
	Features			
	Hybrid	4.5398	4.4670	4.65

Table 5.7: *Cont.*

Scenario	Method	RMSE	MAE	MAPE (%)
5	Proposed	<b>0.0830</b>	<b>0.0757</b>	<b>0.08</b>
	DNN	0.4271	0.3912	0.41
	Direct	4.6395	0.5065	0.53
	Features			
	Hybrid	4.2480	4.1490	4.37
6	Proposed	<b>0.2696</b>	<b>0.2676</b>	<b>0.28</b>
	DNN	22.5040	22.4926	23.58
	Direct	5.5194	2.2520	2.36
	Features			
	Hybrid	4.2111	4.1170	4.30
Average	Proposed	<b>0.0862</b>	<b>0.0732</b>	<b>0.5987</b>
	DNN	5.4385	5.4280	12.0616
	Direct	2.4744	0.5701	2.96725
	Features			
	Hybrid	4.72140	3.2923	3.3591

Table 5.8: Comparative analysis of State of Health (SoH) estimation performance across six scenarios, demonstrating the effect of discharge C-rates, ambient temperatures, and charging protocols on RMSE, MAE, and MAPE metrics.

Scenario	Charge C-Rate	Discharge C-Rate	Ambient Temperature (°C)	RMS	MAE	MAPE (%)
1	~0.75C (CC-CV)	1C	4 to 40	0.002021	0.001711	0.2182
2	~0.75C (CC-CV)	1C	24	0.013571	0.011308	1.5269
3	~0.75C (CC-CV)	2C	4	0.083121	0.028124	1.8028
4	~0.75C (CC-CV)	1C	25	0.1148	0.1147	0.12
5	0.5C (CC)	0.5C (CC)	15	0.2309	0.2308	0.24
6	0.5C (CC)	2C (CC)	25	0.3568	0.3566	0.37

## 5.4 Conclusions

This study demonstrates that the robustness of state-of-health estimation for lithium-ion batteries is significantly influenced by the choice of machine learning architecture, loss function, feature selection, and signal reconstruction technique. The performance of our proposed approach remained stable under low-temperature (4 °C) operation, high discharge rates (2C), severe Gaussian noise (10 dB SNR), and missing data, where DNN and hybrid models lost accuracy. This work also demonstrated that well-engineered features, obtained using domain knowledge, capture relevant information more effectively than a larger quantity of statistics. In this case, our proposed approach, which relies on only five engineered features, outperformed a comparable model that used ten statistical features.

Our proposed closed-form signal reconstruction approach based on modified Tikhonov regularization achieves superior reconstruction quality across various noise levels, compared to the iterative LASSO, moving average filter, and Kalman filter. Our proposed data-driven SoH estimation approach demonstrated high accuracy under noisy conditions, with a low computational cost.

Future research will aim to extend our proposed SoH estimation approach by incorporating additional relevant physical parameters, developing adaptive methods for selecting the regularization parameter ( $\delta$ ), and performing experiments using non-Gaussian noise and extreme operating conditions. Our current results strongly highlight the potential of data-driven methods for achieving accurate and robust SoH estimation within battery management systems (BMS), which are critical for the safe and efficient operation of electric vehicles (EV). Hence, future work will focus on validating our proposed approach under more diverse and variable battery aging conditions, using real-world EVs datasets.

# Chapter 6

## Conclusion and Future Work

### 6.1 Conclusions

This thesis presents two new state estimation methods and two new machine learning approaches for the operation of microgrids equipped with battery units and renewable energy sources. The primary contributions of this research are:

A novel mathematical programming-based state estimation approach is presented for unbalanced three-phase AC microgrids, particularly in scenarios with sparse sensor data. This method demonstrates superior accuracy while maintaining the computational efficiency crucial for real-time operational contexts. The proposed models, based on weighted least squares (WLS) and weighted least absolute value (WLAV) criteria, can be implemented using commercial solvers and can flexibly incorporate various measurement types as constraints, with measurement residuals as explicit optimization variables.

We introduce a novel machine learning state estimation (MLSE) model specifically engineered for unobservable AC microgrids. This innovative approach eliminates the dependency on pseudo-measurements and prior knowledge of measurement error variance, which are typically required by traditional methods. The MLSE model leverages multiple candidate models trained with a limited number of power flow simulations and is enhanced by a proposed Tikhonov regularization operator to effectively handle the heteroscedastic nature of measurements. This method demonstrates robust performance even with noisy measurements and can be adapted to various microgrid configurations.

To address false data injection attacks (FDIAs) in solar power generation systems, we proposed an innovative visual state estimation approach. This method utilizes a modified VGG-16 neural network model to interpret visual data from sky images, mapping these to a virtual solar power generation meter. This creates an additional verification layer, enhancing the integrity and reliability of solar power generation data within Energy Management Systems (EMS).

Finally, in the fourth contribution, we introduced a robust, data-driven State of Health

(SoH) estimation for lithium-ion batteries. Our proposed approach surpassed other models in noisy conditions using limited data.

All the proposed approaches are designed for adaptability and scalability, making them suitable for diverse microgrid configurations operating under various scenarios.

## 6.2 Future Works

Future research directions emerging from this thesis point toward several aspects that can be investigated:

Development of state estimation strategies for unbalanced three-phase systems under contingencies. In this way, important factors that cannot be observed, such as faults in lines and equipment, can be considered in future works.

Developing state estimation strategies considering effects of the battery degradation due to its operation or need of replacement and maintenance of some components of the microgrid.

Considering multi-modal models to create new state estimators using multiple information signals such as images, electrical signals, etc.

Adaptation of state estimation approaches to include additional functionalities such as analysis of power quality issues.

## References

- [1] R. Quadros, J. L. Jucá, J. G. I. Cypriano, R. P. B. D. Silva, L. C. P. D. Silva, and R. G. Bento. Implementation of microgrid on the university campus of UNICAMP - Brazil: Case study. *Journal of Electronics and Advanced Electrical Engineering*, 1(2):21–25, 2021.
- [2] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [3] Pierre-Yves Laffont, Zhile Ren, Xiaofeng Tao, Chao Qian, and James Hays. Transient attributes for high-level understanding and editing of outdoor scenes. *ACM Transactions on Graphics (proceedings of SIGGRAPH)*, 33(4), 2014.
- [4] Nanhua Jiang, Jiawei Zhang, Weiran Jiang, Yao Ren, Jing Lin, Edwin Khoo, and Ziyu Song. Driving behavior-guided battery health monitoring for electric vehicles using extreme learning machine. *Applied Energy*, 364:123122, 2024.
- [5] Lewis Driscoll, Sebastian de la Torre, and Jose Antonio Gomez-Ruiz. Feature-based lithium-ion battery state of health estimation with artificial neural networks. *Journal of Energy Storage*, 50:104584, 2022.
- [6] Junyuan Liang, Hui Liu, and Ning-Cong Xiao. A hybrid approach based on deep neural network and double exponential model for remaining useful life prediction. *Expert Systems with Applications*, 249:123563, 2024.
- [7] Bhaskar Saha and Kai Goebel. Battery data set. *NASA AMES prognostics data repository*, 2007.
- [8] Yuliya Preger, Heather M Barkholtz, Armando Fresquez, Daniel L Campbell, Benjamin W Juba, Jessica Romàn-Kustas, Summer R Ferreira, and Babu Chalamala. Degradation of commercial lithium-ion cells as a function of chemistry and cycling conditions. *Journal of The Electrochemical Society*, 167(12):120532, 2020.
- [9] João Abel Peças Lopes, André Guimarães Madureira, Manuel Matos, Ricardo Jorge Bessa, Vítor Monteiro, João Luiz Afonso, Sérgio F Santos, João PS Catalão, Car-

- los Henggeler Antunes, and Pedro Magalhães. The future of power systems: Challenges, trends, and upcoming paradigms. *Wiley Interdisciplinary Reviews: Energy and Environment*, 9(3):e368, 2020.
- [10] Moses Jeremiah Barasa Kabeyi and Oludolapo Akanni Olanrewaju. Sustainable energy transition for renewable and low carbon grid electricity generation and supply. *Frontiers in Energy research*, 9:743114, 2022.
- [11] C. Proxy Solicitation Services Rendered. Microgrid market forecast to reach \$30.9b by 2027: Navigant. <https://microgridknowledge.com/microgrid-market-navigant/>, 2018. Accessed: 14-Feb-2020.
- [12] H. Bevrani, B. François, and T. Ise. *Microgrid Dynamics and Control*. John Wiley & Sons, 2017.
- [13] A. Primadianto and C.-N. Lu. A review on distribution system state estimation. *IEEE Transactions on Power Systems*, 32(5):3875–3883, September 2017.
- [14] M. M. Rana and L. Li. An overview of distributed microgrid state estimation and control for smart grids. *Sensors*, 15(2):4302–4325, February 2015.
- [15] G. Dong and Z. Chen. Data-driven energy management in a home microgrid based on bayesian optimal algorithm. *IEEE Transactions on Industrial Informatics*, 15(2):869–877, February 2019.
- [16] A. Shafiu, V. Thornley, N. Jenkins, G. Strbac, and A. Maloyd. Control of active networks. In *CIREN 2005 - 18th International Conference and Exhibition on Electricity Distribution*, pages 1–4, June 2005.
- [17] Bang LH Nguyen, Tuyen V Vu, Thomas H Ortmeyer, and Tuan Ngo. Distributed dynamic state estimation for microgrids. In *2020 IEEE Power & Energy Society General Meeting (PESGM)*, pages 1–5. IEEE, 2020.
- [18] C. L. Fortescue. Method of symmetrical co-ordinates applied to the solution of polyphase networks. *Transactions of the American Institute of Electrical Engineers*, 37(2):1027–1140, 1918.
- [19] A. Monticelli. Fast decoupled state estimator. In *State Estimation in Electric Power Systems*, pages 313–342. Springer, 1999.
- [20] Farid Hamzeh Aghdam and Navid Taghizadegan Kalantari. Distributed energy resources and microgrid infrastructure. *Microgrid Architectures, Control and Protection Methods*, pages 69–86, 2020.
- [21] S. Bhela, V. Kekatos, and S. Veeramachaneni. Enhancing observability in distribution grids using smart meter data. *IEEE Transactions on Smart Grid*, 9(6):5953–5961, 2017.

- [22] Marcos Netto, Venkat Krishnan, Yingchen Zhang, and Lamine Mili. Measurement placement in electric power transmission and distribution grids: Review of concepts, methods, and research needs. *IET Generation, Transmission & Distribution*, 16(5):805–838, 2022.
- [23] Byron Alejandro Acuña Acurio, Diana Estefanía Chérrez Barragán, Juan Camilo López, Felipe Grijalva, Juan Carlos Rodríguez, and Luiz Carlos Pereira da Silva. State estimation for unbalanced three-phase ac microgrids based on mathematical programming. In *2023 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–5. IEEE, 2023.
- [24] A. Monticelli. *State Estimation in Electric Power Systems*. Springer, Boston, MA, 1999.
- [25] C.N. Lu, J.H. Teng, and W.-H.E. Liu. Distribution system state estimation. *IEEE Transactions on Power Systems*, 10(1):229–240, 1995.
- [26] Abhinav Kumar Singh, Ravindra Singh, and Bikash C Pal. Stability analysis of networked control in smart grids. *IEEE Transactions on Smart Grid*, 6(1):381–390, 2014.
- [27] David Bakken. *Smart grids: clouds, communications, open source, and automation*. CRC Press, 2014.
- [28] Catalina Gomez-Quiles, Antonio Gomez-Exposito, and Antonio de la Villa Jaen. State estimation for smart distribution substations. *IEEE Transactions on Smart Grid*, 3(2):986–995, 2012.
- [29] Qingyu Yang, Jie Yang, Wei Yu, Dou An, Nan Zhang, and Wei Zhao. On false data-injection attacks against power system state estimation: Modeling and countermeasures. *IEEE Transactions on Parallel and Distributed Systems*, 25(3):717–729, 2013.
- [30] Srinath Kamireddy, Noel N Schulz, and Anurag K Srivastava. Comparison of state estimation algorithms for extreme contingencies. In *2008 40th North American Power Symposium*, pages 1–5. IEEE, 2008.
- [31] Slobodan Pajic and Kevin A Clements. Power system state estimation via globally convergent methods. *IEEE Transactions on Power Systems*, 20(4):1683–1689, 2005.
- [32] R Fourer, D Gay, and B Kernighan. *AMPL: A modeling language for mathematical programming*, brooks/cole–thomson learning. *Pacific Grove*, 2003.
- [33] Lorenz T Biegler and Victor M Zavala. Large-scale nonlinear programming using ipopt: An integrating framework for enterprise-wide dynamic optimization. *Computers & Chemical Engineering*, 33(3):575–582, 2009.

- 
- [34] TP Vishnu, Vidya Viswan, and AM Vipin. Power system state estimation and bad data analysis using weighted least square method. In *Power, Instrumentation, Control and Computing (PICC), 2015 International Conference on*, pages 1–5. IEEE, 2015.
- [35] Pedro P Vergara, Juan C López, Marcos J Rider, and Luiz CP Da Silva. Optimal operation of unbalanced three-phase islanded droop-based microgrids. *IEEE Transactions on Smart Grid*, 10(1):928–940, 2017.
- [36] Jéssica Alice A Silva, Juan Camilo López, Nataly Bañol Arias, Marcos J Rider, and Luiz CP da Silva. An optimal stochastic energy management system for resilient microgrids. *Applied Energy*, 300:117435, 2021.
- [37] Philip E Gill, Walter Murray, and Michael A Saunders. Snopt: An sqp algorithm for large-scale constrained optimization. *SIAM review*, 47(1):99–131, 2005.
- [38] Richard H Byrd, Jorge Nocedal, and Richard A Waltz. K nitro: An integrated package for nonlinear optimization. In *Large-scale nonlinear optimization*, pages 35–59. Springer, 2006.
- [39] A. Abur and A.G. Expósito. *Power System State Estimation: Theory and Implementation*. Power Engineering (Willis). USA: CRC Press, 1 edition, 2004.
- [40] Byron Alejandro Acuña Acurio, Diana Estefanía Chérrez Barragán, Juan Camilo López Amezquita, Marcos J Rider, and Luiz Carlos Pereira Da Silva. Design and implementation of a machine learning state estimation model for unobservable microgrids. *IEEE Access*, 10:123387–123398, 2022.
- [41] Mohamed E El-Hawary. *Advances in Electric Power and Energy: Static State Estimation*. John Wiley & Sons, 2020.
- [42] Fiaz Ahmad, Akhtar Rasool, Emre Ozsoy, Raja Sekar, Asif Sabanovic, and Meltem Elitaş. Distribution system state estimation—a step towards smart grid. *Renewable and Sustainable Energy Reviews*, 81:2659–2671, 2018.
- [43] Kaveh Dehghanpour, Zhaoyu Wang, Jianhui Wang, Yuxuan Yuan, and Fankun Bu. A survey on state estimation techniques and challenges in smart distribution systems. *IEEE Transactions on Smart Grid*, 10(2):2312–2322, 2018.
- [44] A Alimardani, S Zadkhast, J Jatskevich, and E Vaahedi. Using smart meters in state estimation of distribution networks. In *2014 IEEE PES General Meeting/Conference & Exposition*, pages 1–5. IEEE, 2014.
- [45] Alcir Monticelli. Fast decoupled state estimator. In *State Estimation in Electric Power Systems*, pages 313–342. Springer, 1999.
- [46] Siddharth Bhela, Vassilis Kekatos, and Sriharsha Veeramachaneni. Enhancing ob-

- servability in distribution grids using smart meter data. *IEEE Transactions on Smart Grid*, 9(6):5953–5961, 2017.
- [47] Yang Weng, Marija D Ilić, Qiao Li, and Rohit Negi. Convexification of bad data and topology error detection and identification problems in ac electric power systems. *IET Generation, Transmission & Distribution*, 9(16):2760–2767, 2015.
- [48] Anggoro Primadianto and Chan-Nan Lu. A review on distribution system state estimation. *IEEE Transactions on Power Systems*, 32(5):3875–3883, 2016.
- [49] Jing Huang, Vijay Gupta, and Yih-Fang Huang. Electric grid state estimators for distribution systems with microgrids. In *2012 46th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6. IEEE, 2012.
- [50] Ying Hu, Anthony Kuh, Aleksandar Kavcic, and Dora Nakafuji. Real-time state estimation on micro-grids. In *The 2011 International Joint Conference on Neural Networks*, pages 1378–1385. IEEE, 2011.
- [51] Md Rana, Li Li, and Steven Su. Kalman filter based microgrid state estimation and control using the iot with 5g networks. In *2015 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, pages 1–5. IEEE, 2015.
- [52] MM Rana, Wei Xiang, and Eric Wang. H-infinity-based microgrid state estimations using the iot sensors. In *Smart Power Distribution Systems*, pages 285–295. Elsevier, 2019.
- [53] Junbo Zhao, Antonio Gómez-Expósito, Marcos Netto, Lamine Mili, Ali Abur, Vladimir Terzija, Innocent Kamwa, Bikash Pal, Abhinav Kumar Singh, Junjian Qi, et al. Power system dynamic state estimation: Motivations, definitions, methodologies, and future work. *IEEE Transactions on Power Systems*, 34(4):3188–3198, 2019.
- [54] George N Korres, Nikos D Hatziargyriou, and Petros J Katsikas. State estimation in multi-microgrids. *European transactions on electrical power*, 21(2):1178–1199, 2011.
- [55] Junbo Zhao, Marcos Netto, Zhenyu Huang, Shenglong Yu, Antonio Gomez-Expósito, Shaobu Wang, Innocent Kamwa, Shahrokh Akhlaghi, Lamine Mili, Vladimir Terzija, et al. Roles of dynamic state estimation in power system modeling, monitoring and operation. *IEEE Transactions on Power Systems*, 2020.
- [56] Natanael Vieyra, Paul Maya, and Luis M Castro. Dynamic state estimation for microgrid structures. *Electric Power Components and Systems*, 48(3):320–332, 2020.
- [57] Milton Brown Do Coutto Filho and Julio Cesar Stacchini de Souza. Forecasting-aided state estimation—part i: Panorama. *IEEE Transactions on Power Systems*,

- 24(4):1667–1677, 2009.
- [58] Sungyun Choi and AP Sakis Meliopoulos. Effective real-time operation and protection scheme of microgrids using distributed dynamic state estimation. *IEEE Transactions on Power Delivery*, 32(1):504–514, 2016.
- [59] Hany EZ Farag, S Saxena, and A Asif. A robust dynamic state estimation for droop controlled islanded microgrids. *Electric Power Systems Research*, 140:445–455, 2016.
- [60] Eric A Wan and Rudolph Van Der Merwe. The unscented kalman filter for nonlinear estimation. In *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373)*, pages 153–158. IEEE, 2000.
- [61] Urban Kuhar, Gregor Kosec, and Aleš Švigelj. *Observability of Power-Distribution Systems: State-Estimation Techniques and Approaches*. Springer Nature, 2020.
- [62] R Singh, BC Pal, and RA Jabr. Choice of estimator for distribution system state estimation. *IET generation, transmission & distribution*, 3(7):666–678, 2009.
- [63] Nicholas IM Gould, Tyrone Rees, and Jennifer A Scott. Convergence and evaluation-complexity analysis of a regularized tensor-newton method for solving nonlinear least-squares problems. *Computational Optimization and Applications*, 73(1):1–35, 2019.
- [64] Yang Weng, Rohit Negi, Christos Faloutsos, and Marija D Ilić. Robust data-driven state estimation for smart grid. *IEEE Transactions on Smart Grid*, 8(4):1956–1967, 2016.
- [65] Yang Weng, Rohit Negi, and Marija D Ilić. Historical data-driven state estimation for electric power systems. In *2013 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 97–102. IEEE, 2013.
- [66] Kursat Rasim Mestav, Jaime Luengo-Rozas, and Lang Tong. Bayesian state estimation for unobservable distribution systems via deep learning. *IEEE Transactions on Power Systems*, 34(6):4910–4920, 2019.
- [67] Kursat Rasim Mestav and Lang Tong. State estimation in smart distribution systems with deep generative adversary networks. In *2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pages 1–6. IEEE, 2019.
- [68] Christoph Molnar, Giuseppe Casalicchio, and Bernd Bischl. Interpretable machine learning—a brief history, state-of-the-art and challenges. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 417–431. Springer, 2020.

- 
- [69] Roger C. Dugan and Davis Montenegro. *Reference Guide: The Open Distribution System Simulator (OpenDSS)*. EPRI, Jun 2019. Revision: 8.6.1.
- [70] William H Kersting. *Distribution system modeling and analysis*. CRC press, 2017.
- [71] Di Shi, Daniel J Tylavsky, and Naim Logic. An adaptive method for detection and correction of errors in pmu measurements. *IEEE Transactions on Smart Grid*, 3(4):1575–1583, 2012.
- [72] Esmaeil Ghahremani and Innocent Kamwa. Local and wide-area pmu-based decentralized dynamic state estimation in multi-machine power systems. *IEEE Transactions on Power Systems*, 31(1):547–562, 2015.
- [73] Qingqing Huang, Leilai Shao, and Na Li. Dynamic detection of transmission line outages using hidden markov models. *IEEE Transactions on power systems*, 31(3):2026–2033, 2015.
- [74] Ning Zhou. A cross-coherence method for detecting oscillations. *IEEE Transactions on Power Systems*, 31(1):623–631, 2015.
- [75] Le Xie, Yang Chen, and PR Kumar. Dimensionality reduction of synchrophasor data for early event detection: Linearized analysis. *IEEE Transactions on Power Systems*, 29(6):2784–2794, 2014.
- [76] Michael Brown, Milan Biswal, Sukumar Brahma, Satish J Ranade, and Huiping Cao. Characterizing and quantifying noise in pmu data. In *2016 IEEE Power and Energy Society General Meeting (PESGM)*, pages 1–5. IEEE, 2016.
- [77] Raghunath Arnab. *Survey sampling theory and applications*. Academic Press, 2017.
- [78] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.
- [79] Tom Mitchell. *Machine learning*, 1997.
- [80] Kunio Shimizu. *Lognormal distributions: Theory and applications*. Routledge, 2018.
- [81] John Olusegun Ogundare. *Understanding least squares estimation and geomatics data analysis*. John Wiley & Sons, 2018.
- [82] Roger Penrose. On best approximate solutions of linear matrix equations. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 52, pages 17–19. Cambridge University Press, 1956.
- [83] SJ Li and WY Zhang. Hadamard well-posed vector optimization problems. *Journal of Global Optimization*, 46(3):383–393, 2010.

- 
- [84] Laura R Dykes. *New Methods for Solution of Discrete Ill-posed Problems*. PhD thesis, Kent State University, 2016.
- [85] Åke Björck and Lars Eldén. *Methods in numerical algebra for ill-posed problems*. Linköping University, 1979.
- [86] CW Groetsch. The theory of tikhonov regularization for fredholm equations. *104p, Boston Pitman Publication*, 1984.
- [87] Andrey N Tikhonov and Vasiliy Y Arsenin. Solutions of ill-posed problems. *New York*, 1(30):487, 1977.
- [88] Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.
- [89] Kazufumi Ito and Bangti Jin. *Inverse problems: Tikhonov theory and algorithms*, volume 22. World Scientific, 2014.
- [90] Ghadban Khalaf and Ghazi Shukur. *Choosing ridge parameter for regression problems*. Taylor & Francis, 2005.
- [91] D Calvetti, S Morigi, L Reichel, and F Sgallari. Tikhonov regularization and the l-curve for large discrete ill-posed problems. *Journal of computational and applied mathematics*, 123(1-2):423–446, 2000.
- [92] Gilbert Strang. The discrete cosine transform. *SIAM review*, 41(1):135–147, 1999.
- [93] Marco Donatelli and Stefano Serra-Capizzano. Antireflective boundary conditions for deblurring problems. *Journal of Electrical and Computer Engineering*, 2010, 2010.
- [94] Marco Donatelli. Fast transforms for high order boundary conditions in deconvolution problems. *BIT Numerical Mathematics*, 50(3):559–576, 2010.
- [95] Stefano Serra-Capizzano. A note on antireflective boundary conditions and fast deblurring models. *SIAM Journal on Scientific Computing*, 25(4):1307–1325, 2004.
- [96] Richard C Aster, Brian Borchers, and Clifford H Thurber. *Parameter estimation and inverse problems*. Elsevier, 2018.
- [97] Guido Van Rossum and Fred L. Drake. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA, 2009.
- [98] Montgomery DC and Runger GC. *Applied Statistics & Probability for Engineers*. Wiley, 3 edition, 2002.
- [99] Luis F Ugarte, Ji Tuo, and Madson C de Almeida. Assessing approaches to modeling voltage magnitude measurements in state estimators devoted to active distribution

- networks. *Electric Power Systems Research*, 200:107440, 2021.
- [100] Byron Alejandro Acuña Acurio, Diana Estefanía Chérrez Barragán, Juan Camilo López, Felipe Grijalva, Juan Carlos Rodríguez, and Luiz Carlos Pereira da Silva. Visual state estimation for false data injection detection of solar power generation. *Engineering Proceedings*, 47(1):5, 2023.
- [101] Ahmed S Musleh, Guo Chen, and Zhao Yang Dong. A survey on the detection algorithms for false data injection attacks in smart grids. *IEEE Transactions on Smart Grid*, 11(3):2218–2234, 2019.
- [102] Rui Tan, Hoang Hai Nguyen, Eddy YS Foo, David KY Yau, Zbigniew Kalbarczyk, Ravishankar K Iyer, and Hoay Beng Gooi. Modeling and mitigating impact of false data injection attacks on automatic generation control. *IEEE Transactions on Information Forensics and Security*, 12(7):1609–1624, 2017.
- [103] Amir Ameli, Ali Hooshyar, Ehab F El-Saadany, and Amr M Youssef. Attack detection and identification for automatic generation control systems. *IEEE Transactions on Power Systems*, 33(5):4760–4774, 2018.
- [104] Gu Chaojun, Panida Jirutitijaroen, and Mehul Motani. Detecting false data injection attacks in ac state estimation. *IEEE Transactions on Smart Grid*, 6(5):2476–2483, 2015.
- [105] Zhenhua Wang, Haibo He, Zhiqiang Wan, and Yan Sun. Detection of false data injection attacks in ac state estimation using phasor measurements. *IEEE Transactions on Smart Grid*, 2020.
- [106] Kebina Manandhar, Xiaojun Cao, Fei Hu, and Yao Liu. Detection of faults and attacks including false data injection attack in smart grid using kalman filter. *IEEE transactions on control of network systems*, 1(4):370–379, 2014.
- [107] Sen Tan, Peilin Xie, Josep M Guerrero, and Juan C Vasquez. False data injection cyber-attacks detection for multiple dc microgrid clusters. *Applied Energy*, 310:118425, 2022.
- [108] Muhammad Akbar Husnoo, Adnan Anwar, Nasser Hosseinzadeh, Shama Naz Islam, Abdun Naser Mahmood, and Robin Doss. False data injection threats in active distribution systems: A comprehensive survey. *Future Generation Computer Systems*, 2022.
- [109] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.
- [110] Yuhao Nie, Xiatong Li, Andea Scott, Yuchi Sun, Vignesh Venugopal, and Adam Brandt. Skipp’d: A sky images and photovoltaic power generation dataset for

- short-term solar forecasting. *Solar Energy*, 255:171–179, 2023.
- [111] Byron Alejandro Acuña Acurio, Diana Estefanía Chérrez Barragán, Juan Carlos Rodríguez, Felipe Grijalva, and Luiz Carlos Pereira da Silva. Robust data-driven state of health estimation of lithium-ion batteries based on reconstructed signals. *Energies*, 18(10), 2025.
- [112] Peter Greim, AA Solomon, and Christian Breyer. Assessment of lithium criticality in the global energy transition and addressing policy gaps in transportation. *Nature communications*, 11(1):4570, 2020.
- [113] Alexis Lagrange, Miguel de Simón-Martín, Alberto González-Martínez, Stefano Bracco, and Enrique Rosales-Asensio. Sustainable microgrids with energy storage as a means to increase power resilience in critical facilities: An application to a hospital. *International Journal of Electrical Power & Energy Systems*, 119:105865, 2020.
- [114] US Department of Energy DOE. Doe global energy storage database, 2023. Accessed: 2023-09-08.
- [115] Daniel-Ioan Stroe and Erik Schaltz. Lithium-ion battery state-of-health estimation using the incremental capacity analysis technique. *IEEE Transactions on Industry Applications*, 56(1):678–685, 2019.
- [116] Ahmed Yahia Kallel, Viktor Petrychenko, and Olfa Kanoun. State-of-health of lithium battery estimation based on the efficiency of the charge transfer extracted from impedance spectra. *Applied Sciences*, 12(2):885, 2022.
- [117] Tsuyoshi Oji, Yanglin Zhou, Song Ci, Feiyu Kang, Xi Chen, and Xiulan Liu. Data-driven methods for battery soh estimation: Survey and a critical analysis. *Ieee Access*, 9:126903–126916, 2021.
- [118] Jacob John, Ganesh Kudva, and NS Jayalakshmi. Secondary life of electric vehicle batteries: Degradation, state of health estimation using incremental capacity analysis, applications and challenges. *IEEE Access*, 2024.
- [119] Majid Gharebaghi, Omid Rezaei, Changyao Li, Zhanle Wang, and Yili Tang. A survey on using second-life batteries in stationary energy storage applications. *Energies*, 18(1):42, 2024.
- [120] Jianyu Zhang and Kang Li. State-of-health estimation for lithium-ion batteries in hybrid electric vehicles—a review. *Energies*, 17(22):5753, 2024.
- [121] Kailong Liu, Yujie Wang, and Xin Lai. *Data science-based full-lifespan management of lithium-ion battery: manufacturing, operation and reutilization*. Springer Nature, 2022.

- 
- [122] Jinpeng Tian, Rui Xiong, and Weixiang Shen. A review on state of health estimation for lithium ion batteries in photovoltaic systems. *ETransportation*, 2:100028, 2019.
- [123] Jörn Wilhelm, Stefan Seidlmayer, Peter Keil, Jörg Schuster, Armin Kriele, Ralph Gilles, and Andreas Jossen. Cycling capacity recovery effect: A coulombic efficiency and post-mortem study. *Journal of Power Sources*, 365:327–338, 2017.
- [124] Ehab HE Bayoumi, Michele De Santis, and Hilmy Awad. A brief overview of modeling estimation of state of health for an electric vehicle’s li-ion batteries. *World Electric Vehicle Journal*, 16(2):73, 2025.
- [125] Malin Andersson, Moritz Streb, Jing Ying Ko, Verena Löfqvist Klass, Matilda Klett, Henrik Ekström, Mikael Johansson, and Göran Lindbergh. Parametrization of physics-based battery models from input–output data: A review of methodology and current research. *Journal of Power Sources*, 521:230859, 2022.
- [126] Pietro Iurilli, Claudio Brivio, Rafael E Carrillo, and Vanessa Wood. Physics-based soh estimation for li-ion cells. *Batteries*, 8(11):204, 2022.
- [127] Xinwei Sun, Yang Zhang, Yongcheng Zhang, Licheng Wang, and Kai Wang. Summary of health-state estimation of lithium-ion batteries based on electrochemical impedance spectroscopy. *Energies*, 16(15):5682, 2023.
- [128] Manh-Kien Tran, Manoj Mathew, Stefan Janhunen, Satyam Panchal, Kaamran Raahemifar, Roydon Fraser, and Michael Fowler. A comprehensive equivalent circuit model for lithium-ion batteries, incorporating the effects of state of health, state of charge, and temperature on model parameters. *Journal of Energy Storage*, 43:103252, 2021.
- [129] Humberto Velasco-Arellano, Néstor Castillo-Magallanes, Nancy Visairo-Cruz, Ciro Alberto Núñez-Gutiérrez, and Isabel Lázaro. Parametric correlation analysis between equivalent electric circuit model and mechanistic model interpretation for battery internal aging. *World Electric Vehicle Journal*, 15(7):291, 2024.
- [130] Jun Lu, Tianpin Wu, and Khalil Amine. State-of-the-art characterization techniques for advanced lithium-ion batteries. *Nature Energy*, 2(3):1–13, 2017.
- [131] Zhong Ren and Changqing Du. A review of machine learning state-of-charge and state-of-health estimation algorithms for lithium-ion batteries. *Energy Reports*, 9:2993–3021, 2023.
- [132] Ming Zhang, Dongfang Yang, Jiaxuan Du, Hanlei Sun, Liwei Li, Licheng Wang, and Kai Wang. A review of soh prediction of li-ion batteries based on data-driven algorithms. *Energies*, 16(7):3167, 2023.
- [133] Jianqiang Gong, Bin Xu, Fanghua Chen, and Gang Zhou. Predictive modeling for

- electric vehicle battery state of health: A comprehensive literature review. *Energies*, 18(2):337, 2025.
- [134] Gulzat Nuroldayeva, Yerkin Serik, Desmond Adair, Berik Uzakbaiuly, and Zhumabay Bakenov. State of health estimation methods for lithium-ion batteries. *International Journal of Energy Research*, 2023(1):4297545, 2023.
- [135] Khaled Sidahmed Sidahmed Alamin, Francesco Daghero, Giovanni Pollo, Daniele Jahier Pagliari, Yukai Chen, Enrico Macii, Massimo Poncino, and Sara Vinco. Model-driven dataset generation for data-driven battery soh models. In *2023 IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED)*, pages 1–6. IEEE, 2023.
- [136] Safieh Bamati and Hicham Chaoui. Developing an online data-driven state of health estimation of lithium-ion batteries under random sensor measurement unavailability. *IEEE Transactions on Transportation Electrification*, 9(1):1128–1141, 2022.
- [137] Jinyu Wang, Caiping Zhang, Xiangfeng Meng, Linjing Zhang, Xu Li, and Weige Zhang. A novel feature engineering-based soh estimation method for lithium-ion battery with downgraded laboratory data. *Batteries*, 10(4):139, 2024.
- [138] Zhao Zhang, Runrun Zhang, Xin Liu, Chaolong Zhang, Gengzhi Sun, Yujie Zhou, Zhong Yang, Xuming Liu, Shi Chen, Xinyu Dong, et al. Advanced state-of-health estimation for lithium-ion batteries using multi-feature fusion and kan-lstm hybrid model. *Batteries*, 10(12):433, 2024.
- [139] Xin Lu, Jing Qiu, Gang Lei, and Jianguo Zhu. Degradation mode knowledge transfer method for lfp batteries. *IEEE Transactions on Transportation Electrification*, 9(1):1142–1152, 2022.
- [140] Xin Lu, Jing Qiu, Gang Lei, and Jianguo Zhu. State of health estimation of lithium iron phosphate batteries based on degradation knowledge transfer learning. *IEEE Transactions on Transportation Electrification*, 9(3):4692–4703, 2023.
- [141] Raid Mohsen Alhazmi. State of health prediction in electric vehicle batteries using a deep learning model. *World Electric Vehicle Journal*, 15(9):385, 2024.
- [142] Phattara Khumprom and Nita Yodo. A data-driven predictive prognostic model for lithium-ion batteries based on a deep learning algorithm. *Energies*, 12(4):660, 2019.
- [143] Zhiquan Cui, Chunhui Wang, Xuhong Gao, and Shushan Tian. State of health estimation for lithium-ion battery based on the coupling-loop nonlinear autoregressive with exogenous inputs neural network. *Electrochimica Acta*, 393:139047, 2021.
- [144] Yaxiang Fan, Fei Xiao, Chaoran Li, Guorun Yang, and Xin Tang. A novel deep learning framework for state of health estimation of lithium-ion battery. *Journal of*

- Energy Storage*, 32:101741, 2020.
- [145] Danhua Zhou, Zhanying Li, Jiali Zhu, Haichuan Zhang, and Lin Hou. State of health monitoring and remaining useful life prediction of lithium-ion batteries based on temporal convolutional network. *IEEE Access*, 8:53307–53320, 2020.
- [146] Carlos Antônio Rufino Júnior, Eleonora Riva Sanseverino, Pierluigi Gallo, Murilo Machado Amaral, Daniel Koch, Yash Kotak, Sergej Diel, Gero Walter, Hans-Georg Schweiger, and Hudson Zanin. Unraveling the degradation mechanisms of lithium-ion batteries. *Energies*, 17(14), 2024.
- [147] Yifan Wei, Yuan Yao, Kang Pang, Chaojie Xu, Xuebing Han, Languang Lu, Yalun Li, Yudi Qin, Yuejiu Zheng, Hewu Wang, et al. A comprehensive study of degradation characteristics and mechanisms of commercial li (nimnco) o2 ev batteries under vehicle-to-grid (v2g) services. *Batteries*, 8(10):188, 2022.
- [148] Mehmet C Yagci, Thomas Feldmann, Elmar Bollin, Michael Schmidt, and Wolfgang G Bessler. Aging characteristics of stationary lithium-ion battery systems with serial and parallel cell configurations. *Energies*, 15(11):3922, 2022.
- [149] Jacqueline S Edge, Simon O’Kane, Ryan Prosser, Niall D Kirkaldy, Anisha N Patel, Alastair Hales, Abir Ghosh, Weilong Ai, Jingyi Chen, Jiang Yang, et al. Lithium ion battery degradation: what you need to know. *Physical Chemistry Chemical Physics*, 23(14):8200–8221, 2021.
- [150] Xiaoyu Yang, Zhipeng Wang, and Song Xie. Influence of overdischarge depth on the aging and thermal safety of lini0. 5co0. 2mn0. 3o2/graphite cells. *Battery Energy*, page e70008, 2025.
- [151] Shuhui Li and Bao Ke. Study of battery modeling using mathematical and circuit oriented approaches. In *2011 IEEE power and energy society general meeting*, pages 1–8. IEEE, 2011.
- [152] Christoph R Birkl, Matthew R Roberts, Euan McTurk, Peter G Bruce, and David A Howey. Degradation diagnostics for lithium ion cells. *Journal of Power Sources*, 341:373–386, 2017.
- [153] Wei Shen, Ning Wang, Jun Zhang, Feng Wang, and Guangxu Zhang. Heat generation and degradation mechanism of lithium-ion batteries during high-temperature aging. *ACS omega*, 7(49):44733–44742, 2022.
- [154] Rui Huang, Yidan Xu, Qichao Wu, Junxuan Chen, Fenfang Chen, and Xiaoli Yu. Simulation study on heat generation characteristics of lithium-ion battery aging process. *Electronics*, 12(6):1444, 2023.
- [155] Qi Wang, Yue Ma, Kun Zhao, and Yingjie Tian. A comprehensive survey of loss

- functions in machine learning. *Annals of Data Science*, pages 1–26, 2020.
- [156] Olvi L Mangasarian and David R. Musicant. Robust linear and support vector regression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(9):950–955, 2000.
- [157] Peter J Huber. Robust estimation of a location parameter. In *Breakthroughs in statistics: Methodology and distribution*, pages 492–518. Springer, 1992.
- [158] Shuai Ma, Modi Jiang, Peng Tao, Chengyi Song, Jianbo Wu, Jun Wang, Tao Deng, and Wen Shang. Temperature effect and thermal impact in lithium-ion batteries: A review. *Progress in Natural Science: Materials International*, 28(6):653–666, 2018.
- [159] Battery Archive. Battery archive. <https://www.batteryarchive.org/>, 2025. Accessed on 6 March 2025.
- [160] Molla Shahadat Hossain Lipu, Abdullah Al Mamun, Shaheer Ansari, Md Sazal Miah, Kamrul Hasan, Sheikh T Meraj, Maher GM Abdolrasol, Tuhibur Rahman, Md Hasan Maruf, Mahidur R Sarker, et al. Battery management, key technologies, methods, issues, and future trends of electric vehicles: A pathway toward achieving sustainable development goals. *Batteries*, 8(9):119, 2022.
- [161] Python Software Foundation. *Python Programming Language*. Python Software Foundation, Wilmington, Delaware, USA, 1991–. <https://www.python.org/>.
- [162] Mattin Lucu, Egoitz Martinez-Laserna, Iñigo Gandiaga, and Haritza Camblong. A critical review on self-adaptive li-ion battery ageing models. *Journal of Power Sources*, 401:85–101, 2018.
- [163] Shlomo Engelberg. Implementing moving average filters using recursion [tips & tricks]. *IEEE Signal Processing Magazine*, 40(7):78–80, 2023.
- [164] Youngjoo Kim and Hyochoong Bang. Introduction to kalman filter and its applications. In *Introduction and implementations of the Kalman filter*. IntechOpen, 2018.
- [165] Mingqiang Lin, Xianping Zeng, and Ji Wu. State of health estimation of lithium-ion battery based on an adaptive tunable hybrid radial basis function network. *Journal of Power Sources*, 504:230063, 2021.
- [166] Li-Ming Deng, Yu-Cheng Hsu, and Han-Xiong Li. An improved model for remaining useful life prediction on capacity degradation and regeneration of lithium-ion battery. In *Annual Conference of the PHM Society*, volume 9, 2017.
- [167] Diana Estefanía Chérrez Barragán, Byron Alejandro Acuña Acurio, Juan López, Hugo Morais, Cindy P Guzman, and Luiz Silva. Day-ahead photovoltaic power forecasting with limited data. In *2024 IEEE URUCON*, pages 1–5. IEEE, 2024.

- 
- [168] Diana Estefanía Chérrez Barragán, Byron Alejandro Acuña Acurio, Juan Carlos Rodríguez, Felipe Grijalva, Tisciane Perpétuo e Oliveira, and Luiz Carlos Pereira Da Silva. Towards sustainable development based on university microgrids. In *2024 IEEE Eighth Ecuador Technical Chapters Meeting (ETCM)*, pages 1–6. IEEE, 2024.
- [169] Juan Carlos Rodriguez, Felipe Grijalva, Marcelo García, Diana Estefanía Chérrez Barragán, Byron Alejandro Acuña Acurio, and Henry Carvajal. Wireless communication technologies for smart grid distribution networks. *Engineering Proceedings*, 47(1):7, 2023.
- [170] Diana Estefanía Chérrez Barragán, Byron Alejandro Acuña Acurio, Juan Camilo López, Felipe Grijalva, Juan Carlos Rodríguez, and Luiz Carlos Pereira da Silva. An energy management system for a residential microgrid using convex optimization. In *2022 IEEE Sixth Ecuador Technical Chapters Meeting (ETCM)*, pages 1–5. IEEE, 2022.

# Appendix A

## Permission Grants

5/8/25, 10:11 AM

Rightslink® by Copyright Clearance Center


[Sign in/Register](#)


### State Estimation for Unbalanced Three-Phase AC Microgrids Based on Mathematical Programming

Conference Proceedings:  
2023 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)  
Author: Byron Alejandro Acuña Acurio  
Publisher: IEEE  
Date: 16 January 2023

Copyright © 2023, IEEE

### Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

*Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:*

- 1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
- 2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
- 3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

*Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:*

- 1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
- 2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis online.
- 3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to [http://www.ieee.org/publications\\_standards/publications/rights/rights\\_link.html](http://www.ieee.org/publications_standards/publications/rights/rights_link.html) to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

[BACK](#)
[CLOSE WINDOW](#)

© 2025 Copyright - All Rights Reserved | [Copyright Clearance Center, Inc.](#) | [Privacy statement](#) | [Data Security and Privacy](#)  
 | [For California Residents](#) | [Terms and Conditions](#) Comments? We would like to hear from you. E-mail us at [customer-care@copyright.com](mailto:customer-care@copyright.com)



# Appendix **B**

## List of Publications

**Byron Alejandro Acuña Acurio**, Diana Estefanía Chérrez Barragán, Juan Camilo López, Felipe Grijalva, Juan Carlos Rodríguez and Luiz Carlos Pereira da Silva, "State Estimation for Unbalanced Three-Phase AC Microgrids Based on Mathematical Programming" [23] in IEEE ISGT 2023,  
doi: <https://doi.org/10.1109/ISGT51731.2023.10066353>

**Byron Alejandro Acuña Acurio**, Diana Estefanía Chérrez Barragán, Juan Camilo López, Marcos J. Rider and Luiz Carlos Pereira da Silva, "Design and Implementation of a Machine Learning State Estimation Model for Unobservable Microgrids" [40] in IEEE ACCESS 2022,  
doi: <https://doi.org/10.1109/ACCESS.2022.3224758>

**Byron Alejandro Acuña Acurio**, Diana Estefanía Chérrez Barragán, Juan Camilo López, Felipe Grijalva, Juan Carlos Rodríguez and Luiz Carlos Pereira da Silva, "Visual State Estimation for False Data Injection Detection of Solar Power Generation" [100] in Journal of Engineering Proceedings 2024,  
doi: <https://doi.org/10.3390/engproc2023047005>

**Byron Alejandro Acuña Acurio**, Diana Estefanía Chérrez Barragán, Juan Carlos Rodríguez, Felipe Grijalva and Luiz Carlos Pereira da Silva, "Robust Data-Driven State of Health Estimation of Lithium-Ion Batteries Based on Reconstructed Signals" [111] in Energies 2025,  
doi: <https://doi.org/10.3390/en18102459>

Diana Estefanía Chérrez Barragán, **Byron Alejandro Acuña Acurio**, Juan Camilo López, Hugo Morais, Cindy P. Guzman, and Luiz Carlos Pereira Da Silva. "Day-Ahead Photovoltaic Power Forecasting with Limited Data" [167],  
doi: <https://doi.org/10.1109/URUCON63440.2024.10850063>

Diana Estefanía Chérrez Barragán, **Byron Alejandro Acuña Acurio**, Juan Carlos Rodríguez, Felipe Grijalva, Tisciane Perpétuo e Oliveira, Luiz Carlos Pereira Da Silva. "Towards Sustainable Development Based on University Microgrids" [168], doi: <https://doi.org/10.1109/ETCM63562.2024.10746179>

Juan Carlos Rodriguez, Felipe Grijalva, Marcelo García, Diana Estefanía Chérrez Barragán, **Byron Alejandro Acuña Acurio**, Henry Carvajal. "Wireless Communication Technologies for Smart Grid Distribution Networks" [169], doi: <https://doi.org/10.3390/engproc2023047007>

Diana Estefanía Chérrez Barragán, **Byron Alejandro Acuña Acurio**, Juan Camilo López, Felipe Grijalva, Juan Carlos Rodríguez, and Luiz Carlos Pereira Da Silva. "An Energy Management System for a Residential Microgrid using Convex Optimization" [170], doi: <https://doi.org/10.1109/ETCM56276.2022.9935726>