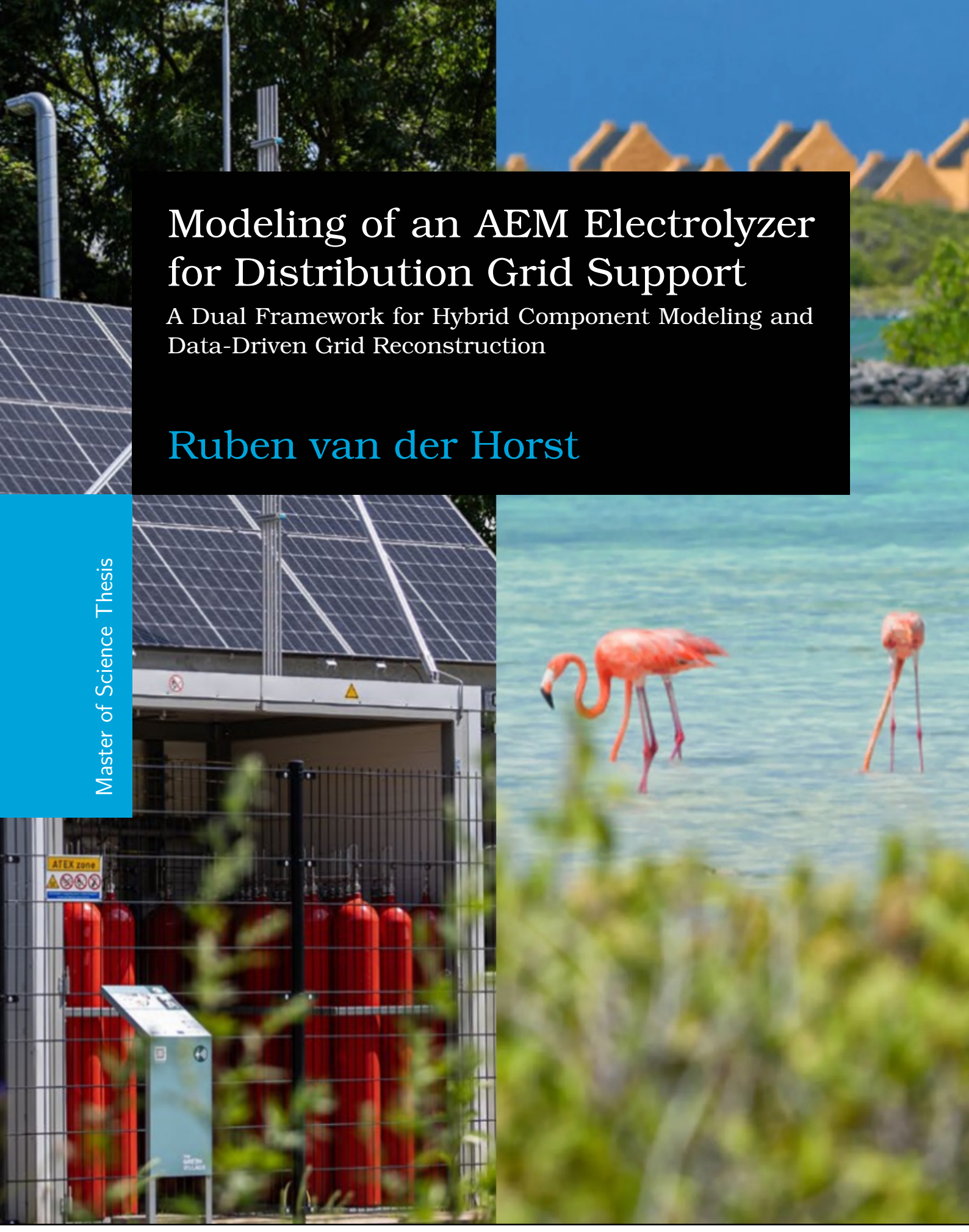

Modeling of an AEM Electrolyzer for Distribution Grid Support

A Dual Framework for Hybrid Component Modeling and Data-Driven Grid Reconstruction

Ruben van der Horst

Master of Science Thesis

TU Delft Delft University of Technology

Delft Center for Systems and Control

Modeling of an AEM Electrolyzer for Distribution Grid Support

A Dual Framework for Hybrid Component Modeling and Data-Driven Grid Reconstruction

Master of Science Thesis

For the degree of Master of Science in Systems and Control at Delft University of Technology

Ruben van der Horst

November 11, 2025

Faculty of Mechanical Engineering (ME) · Delft University of Technology

The work in this thesis was supported by NTCS Energy, WEB Bonaire and The Green Village.

"But still try for who knows what is possible!"
— *Michael Faraday*

Contents

Abstract

The global energy transition towards Renewable Energy Sources (RES) introduces significant challenges to grid stability due to the inherent intermittency of sources like solar and wind. Hybrid Energy Storage (HES) systems, particularly those integrating Power-to-Hydrogen-to-Power (P2H2P) technologies, are critical for mitigating these issues. However, the effective planning and control of these systems depend on accurate models of both the HES components and the Low Voltage (LV) distribution grids they support. This thesis addresses this dual modeling challenge, which is often hindered by complex component dynamics and limited grid data.

This research develops and validates two distinct frameworks. First, a control-oriented, hybrid dynamic model for an Anion Exchange Membrane (AEM) electrolyzer is proposed. This model integrates a physics-based Equivalent Circuit Model (ECM) to capture primary electrochemical dynamics with a shallow neural network trained to correct for residual, unmodeled thermal dynamics. Second, a complete, data-driven pipeline is developed to construct "digital twins" of LV distribution grids from limited static data. This pipeline uses an Integer Linear Programming (ILP) algorithm to reconstruct the most probable network topology, which is then parameterized for power flow analysis.

The hybrid electrolyzer model was validated using operational data from a commercial Enapter EL 2.1 unit, demonstrating high fidelity with a 92.03 % fit, drastically outperforming the 26.00% fit of the standalone ECM. The grid modeling pipeline was validated in a case study of a 160kVA substation in Bonaire. Despite incomplete GIS data, the resulting model accurately simulated aggregate load profiles, achieving a low Root Mean Square Error (RMSE) of approximately 6% for both active and reactive power when compared to real-world measurements.

In synthesis, this thesis delivers a robust and validated toolset for analyzing HES-grid interaction: a high-fidelity, computationally efficient model of an AEM electrolyzer and a viable method for accurately modeling limited-data LV grids. These components enable advanced co-simulation and control design to study the role of AEM electrolyzers in providing distribution grid support.

Chapter 1

Introduction

1-1 The challenge of renewable energy sources

The global energy landscape is currently undergoing a profound transformation, driven by the urgent need to mitigate climate change by transitioning away from conventional fossil fuels [1]. This shift is characterized by an increasing reliance on Renewable Energy Sources (RES), such as solar and wind power. However, the inherent intermittency and variability of these sources pose a significant challenge to the stability and reliability of modern electrical grids. This variability creates a potential mismatch between variable power generation and fluctuating consumer demand, which can threaten grid stability and security of supply [2].

To ensure a stable and reliable power system in the face of high RES penetration, advanced technological solutions are essential. Energy Storage Systems (ESS) have emerged as a critical enabling technology to address this challenge of intermittency. By absorbing surplus energy during periods of high generation and dispatching it when demand exceeds supply, ESS are indispensable for the large-scale integration of renewables [3].

1-2 Modeling of energy storage and distribution systems

Among the various storage solutions, Hybrid Energy Storage (HES) systems are particularly promising. These systems combine two or more storage technologies to leverage their complementary strengths, such as pairing the rapid response of batteries with the long-duration, high-capacity storage of a Power-to-Hydrogen-to-Power (P2H2P) cycle [4]. The integration of these advanced, decentralized assets, however, introduces significant new complexities for grid management.

The effective planning, placement, and control of these HES systems are entirely dependent on the availability of reliable and accurate models. This thesis addresses two primary and interconnected modeling challenges:

1. **Modeling the HES Components:** To integrate HES systems with intermittent renewables, the components themselves, particularly electrolyzers, must operate under highly dynamic loads. This drives the need for accurate, efficient, and control-oriented dynamic models that can capture complex behaviors, such as thermal effects, under variable conditions to estimate the best possible dispatching strategy. Creating such models faces challenges due to the inherent complexity of electrochemical systems [5] and the fact that measurement data is often scarce or noisy. These models are essential for advanced control strategies like Model Predictive Control (MPC) [6–8].
2. **Modeling the Distribution Grid:** The HES systems must be integrated into the existing power grid. Therefore, a reliable model of the distribution grid itself is paramount. Distribution Service Operators (DSO) rely on these models to perform advanced simulations, assess hosting capacity for new Distributed Energy Resources (DERs) [7], and optimize storage dispatch to prevent grid congestion [9, 10]. This task is especially challenging in many Low Voltage (LV) networks, which often suffer from limited data, incomplete documentation, and a lack of advanced metering infrastructure.

This research confronts both of these challenges by developing and validating advanced modeling methodologies for both the HES components and the distribution grids they connect to.

1-3 Research objectives and scope

Based on the challenges identified, the primary identified research gap is:

A lack of accurate, computationally tractable, and physically interpretable models for HES system components (especially emerging tech like AEM electrolyzers) and their interaction with the grid.

There are clear opportunities do utilize AEM electrolyzers in HES systems in an optimal fashion with modern, model driven control strategies. These strategies have been proposed for simple linear and piecewise affine models, but have not been combined with accurate grid models.

To generate these models, the two following secondary objectives are defined:

1. To develop and validate a control-oriented, hybrid dynamic model for an Anion Exchange Membrane (AEM) electrolyzer. This model aims to balance physical interpretability with high predictive accuracy, specifically by capturing the complex, unmodeled thermal dynamics that arise from variable load operation.
2. To develop and validate a complete, data-driven pipeline for modeling LV distribution grids where data sources are limited or incomplete. The objective is to create a reliable grid model suitable for power flow analysis and HES integration studies, based on static GIS and asset data.

The scope of this work is centered on the development, application, and validation of these modeling frameworks. The electrolyzer model (Chapter 2) is validated using real-world

operational data from a commercial Enapter EL 2.1 AEM electrolyzer installed at the 24/7 Energy Hub at The Green Village, TU Delft. The distribution grid modeling methodology (Chapter 3) is applied and validated in a case study of a 160kVA secondary substation in Bonaire, using utility data from WEB Bonaire.

1-4 Research methodology and thesis outline

This thesis addresses its research objectives in distinct chapters, each structured as a self-contained paper. While interconnected, every chapter presents its own methodology, literature review, and results, allowing it to be read independently.

Chapter 2 proposes a hybrid modeling framework for the AEM electrolyzer. It integrates a physics-based Equivalent Circuit Models (ECM), capturing dominant electrochemical dynamics, with a shallow neural network. The network is trained to correct residual errors identified as complex, unmodeled temperature-dependent dynamics, blending interpretability with computational efficiency.

Chapter 3 details a data-driven pipeline to model LV distribution grids. First, it reconstructs the most probable network topology from GIS and customer data using an Integer Linear Programming (ILP) algorithm [10]. Next, grid assets (lines, transformers, loads) are parameterized within the `power-grid-model` library [11]. The resulting model is validated by comparing its simulated aggregate load profile against real-world power quality meter data.

Chapter 4 summarizes the findings of each research chapter, presents the overall conclusions of this thesis, and provides recommendations for future work.

Chapter 2

Control-Oriented Hybrid Modeling of AEM Electrolyzer via Residual Error Correction

2-1 Introduction

Climate change and global warming, driven primarily by human activities that result in excessive emissions of greenhouse gases such as carbon dioxide, have emerged as pressing global challenges [12]. Accordingly, the Paris Agreement has set ambitious targets to limit the rise in global temperatures to 1.5 °C above pre-industrial levels and significantly reduce greenhouse gas emissions, with a particular emphasis on decarbonizing the energy and industrial sectors [13]. Transitioning away from conventional fossil fuel-based energy production and industrial processes necessitates the extensive deployment of renewable energy sources, such as solar and wind power. However, the inherent intermittency and variability of renewable energy generation pose significant challenges to ensuring the stability and reliability of modern power systems. Accordingly, advanced technological solutions are essential to provide effective renewable integration [14]. To address these challenges, robust energy storage approaches are crucial for maintaining grid stability and securing a continuous power supply. Hybrid Energy Storage (HES) systems particularly offer a promising solution. More precisely, HES systems allow to combine two or more energy storage technologies and leverage their complementary strengths to effectively manage renewable energy fluctuations [15, 16].

Within HES systems, green hydrogen, produced via water electrolysis powered by renewables, is emerging as a critical energy carrier. The electrolyzer is a central component. Among various electrolyzer technologies, the Anion Exchange Membrane (AEM) electrolyzer is gaining traction and widely adopted [17]. To integrate AEM electrolyzers with intermittent renewable energy sources, they must operate under highly dynamic loads, causing fluctuations in current, temperature, and pressure. Understanding these effects on performance and degradation is key to reliable design [18, 19], driving the need for accurate, efficient dynamic models.

Efficient use of HES in the energy transition requires advanced control and optimization techniques, such as Model Predictive Control (MPC), which are predominantly model-based, and thus, making the development of accurate component models essential. However, creating such models faces significant challenges, including the inherent complexity of the electrochemical systems [5] and the fact that measurement data is often noisy, scarce, or not sufficiently informative. Furthermore, the resulting optimization and control needs to remain tractable, particularly for decision-making [8, 20–22].

The modeling field for PEM and AEM electrolyzers is notably deep, evolving from simple mathematical equivalents to sophisticated data-driven and hybrid frameworks [23–26]. Within Equivalent Circuit Models (ECM)s, the complexity of the Resistance-Capacitance (RC) configuration directly dictates the captured dynamic range [27–29]. The simplest structure is the zero-order model, which uses only a series resistor to represent instantaneous ohmic losses, failing to capture any transient response [27]. The first-order model adds a single parallel RC branch to capture the fast, namely in millisecond scale, activation dynamics, improving short-term transient accuracy [28,30]. For better representation of both fast and slow dynamic phenomena, possibly in second scale, the double-branch RC model incorporates a second RC branch, often dedicated to the slower concentration and diffusion losses, providing an improved suitability for control-oriented applications [31, 32]. Higher-order models, while offering incrementally higher fidelity, by adding more RC pairs or inclusin of partial differential equations of electrochemical phenomena, typically suffer from increased computational cost and parameter identifiability challenges [33–35]. One the other hand, data-driven approaches, such as neural networks [36], have proven effective for capturing complex, non-linear dynamics, for instance, in the direct operational data-driven dynamic voltage prediction of commercial alkaline water electrolyzers [37]. Moving further into advanced machine learning, Physics-Informed Neural Networks (PINN)s represent a critical development, as they successfully merge deep learning with fundamental electrochemical principles. PINN ensure physical consistency and interpretability, with recent applications demonstrating their power in predicting critical variables like temperature fluctuations in PEM cells [38]. This hybrid strategy, often implemented using residual learning [39], integrates physics-based models with neural networks to specifically correct for unmodeled dynamics and errors [40]. Techniques like the Dynamic Residual Learning approach [41] further improve the fidelity of physics-constrained neural networks, offering more accurate models.

The literature on electrolyzer modeling reveals a fundamental trade-off between identifiability, predictive accuracy and computational complexity in the sense of being suitable for online decision-making and optimization-based control strategies such as MPC. High-fidelity, multi-dimensional models for PEM and AEM electrolyzers provide deep physical insights but are often considerably computationally intensive for real-time control applications or with intractable parameter estimation [42,43]. Conversely, while computationally efficient, simplified approaches like ECMs, often fail to capture critical non-linear dynamics, particularly thermal effects crucial for performance prediction under variable loads [44,45]. Their internal structure, such as RC configuration, is deliberately simplified, e.g., by modeling anodic and cathodic dynamics with a single loop, to prioritize this computational speed [45]. Furthermore, Similarly, methods based on extensive large-scale Neural Networks (NNs) face significant challenges in real-world advanced optimization control strategies; their inherent complexity presents a major hurdle for efficient execution within optimization solvers, and providing technical guarantees of stability or performance is generally not feasible for such black-box models.

Figure 2-1: The 24/7 energy hub located at The Green Village, TU Delft campus.

Consequently, a significant gap exists for modeling frameworks, particularly for emerging AEM technologies, that can effectively balance prediction accuracy, computational efficiency, and physical interpretability [46].

Motivated by the aforementioned research gap, the primary contribution of this work is a control-oriented hybrid modeling framework designed to systematically balance physical interpretability, predictive accuracy, and computational efficiency for real-time control. Unlike traditional methods that embed physics into the loss function, a deliberate two-stage estimation strategy is introduced: an interpretable, physics-based RC model is first identified to capture dominant dynamics. A computationally efficient, shallow neural network is then trained to exclusively correct the remaining residual, which is identified as stemming from complex, unmodeled temperature-dependent dynamics. This hybrid model design, for instance by using bounded activation functions, can also ensure system-theoretic properties by quantifying and controlling the deviation from the physics-based trend. The approach was successfully validated using real-world operational data from a commercial AEM electrolyzer at The Green Village, capturing these thermal effects without elaborate multiphysics modeling. The resulting framework provides a high-fidelity, tractable solution that decisively outperforms baseline physics-only and standard data-driven benchmarks

2-2 The hybrid modeling framework for electrochemical systems

Electrochemical energy systems, such as electrolyzers, fuel cells, and lithium-ion batteries, exhibit nonlinear and multi-scale dynamics due to the intricate coupling of electrochemical, thermal, and transport phenomena, including charge transfer kinetics, ionic and electronic conduction, gas-liquid phase interactions, heat generation, and faradaic reactions [5]. The interplay among the mentioned factors majorly governs the transient response and also the

long-term steady-state characterizations of the system, particularly under dynamic loading conditions [5]. Consequently, to accurately capture and predict their behavior, dynamic systems modeling frameworks are required that effectively encode complex interdependencies across multiple phenomena and timescales, and are also suitable for control-oriented objectives. Accordingly, given an appropriate sampling interval, one may employ a discrete-time state-space representation to describe the system dynamics as

$$x_{k+1} = f(x_k, u_k, d_k), \tag{2-1}$$

where $k \in \mathbb{Z}$ is the discrete index of time, $x_k \in \mathbb{R}^{n_x}$ denotes the vector of internal system states, $u_k \in \mathbb{R}^{n_u}$ represents the control inputs, e.g., current or voltage, $d_k \in \mathbb{R}^{n_d}$ accounts for the vector of internal or external disturbances such as temperature fluctuations or load changes, and, $f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_d} \to \mathbb{R}^{n_x}$ is the mathematical expression for the dynamics of the systems. Alternatively, one may employ an input-state-output or an input-output representation for the system. Nevertheless, the core arguments of our discussion remain unaffected. Therefore, for streamlined exposition, the primary focus is maintained on the input-state settings introduced above.

Let $\mathscr{M} \subseteq \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_d}$ denote the set of admissible vectors of state variables, inputs, and disturbances, and suppose a set of $n_{\mathscr{D}}$ measurement data points in $\mathscr{M}$, denoted by $\mathscr{D}$, is provided as

$$\mathscr{D} = \{(x_k, u_k, d_k) \in \mathscr{M} \,|\, k = 1, \ldots, n_{\mathscr{D}}\}. \tag{2-2}$$

The main objective is to utilize dataset $\mathscr{D}$ to identify a model for the system [47], i.e., to approximate the underlying system law f introduced in (2-1).

Following first principles and the physics of the system, one can employ a *white-box* modeling approach. More precisely, considering the coupled electrochemical and physical processes that govern the dynamics of the system, one may develop parametric models with various components describing different essential features of the system [5]. Thus, a class of physics-based models is obtained as

$$\mathscr{F}^{(\text{phys})} = \left\{ f_\theta^{(\text{phys})} : \mathscr{M} \to \mathbb{R}^{n_x} \,|\, \theta \in \Theta \right\}, \tag{2-3}$$

where $\Theta \subseteq \mathbb{R}^{n_\theta}$ represents the set of admissible parameters. Subsequently, the model identification can be formulated in form of an optimization problem as

$$\min_{\theta \in \Theta} \mathcal{E}_{\mathscr{D}}^{(\text{phys})}(\theta) + \mathcal{R}^{(\text{phys})}(\theta), \tag{2-4}$$

where $\mathcal{E}_{\mathscr{D}}^{(\text{phys})} : \Theta \to \mathbb{R}_{\geq 0}$ is a suitable *empirical loss* indicating model fitting performance given data set $\mathscr{D}$, and $\mathcal{R}^{(\text{phys})} : \Theta \to \mathbb{R}_{\geq 0}$ can be the *regularization* term to avoid overfitting, or to promote specific desired properties for the parameter vectors θ. Note that different forms of $\mathcal{E}_{\mathscr{D}}$ may be employed, including the sum of squared one-step or multi-step ahead prediction errors, or the norm of the simulation error [47]. The elements within the model class $\mathscr{F}^{(\text{phys})}$ are inherently interpretable and can typically capture the qualitative behavior of the system. By employing advanced modeling techniques, one may construct $\mathscr{F}^{(\text{phys})}$ to include high-fidelity models. However, developing such models often demands significant domain expertise. Moreover, estimating their parameters can be computationally intensive, particularly when the available operational data are noisy, limited, or uninformative. In

addition, the parameter estimation problem (2-4) is often highly nonlinear and nonconvex. Conversely, restricting $\mathscr{F}^{(\mathrm{phys})}$ to simplified models, while still preserving interpretability and the ability to reproduce general system trends, often leads to inaccurate representations of the actual dynamics.

One may employ *black-box* modeling frameworks, i.e., purely data-driven schemes. To this end, a suitable parametric model class is considered as

$$\mathscr{F}^{(\mathrm{BB})} = \left\{ f_\xi^{(\mathrm{BB})} : \mathscr{M} \to \mathbb{R}^{n_x} \,|\, \xi \in \Xi \right\}, \tag{2-5}$$

where $\xi \in \mathbb{R}^{n_\xi}$ denotes the parameter vector and Ξ represents the set of feasible parameters. Subsequently, the parameters of the model are estimated using dataset $\mathscr{D}$ through solving an optimization problem formulated similar to (2-4), i.e.,

$$\min_{\xi \in \Xi} \mathcal{E}_{\mathscr{D}}^{(\mathrm{BB})}(\xi) + \mathcal{R}^{(\mathrm{BB})}(\xi), \tag{2-6}$$

where $\mathcal{E}_{\mathscr{D}}^{(\mathrm{BB})} : \Xi \to \mathbb{R}_{\geq 0}$ and $\mathcal{R}^{(\mathrm{BB})} : \Xi \to \mathbb{R}_{\geq 0}$ are defined analogously to (2-4). The elements of $\mathscr{F}^{(\mathrm{BB})}$ can be characterized through various forms. For example, $f_\xi^{(\mathrm{BB})}$ may be described as a linear combination of a given set of functions $\varphi_1, \ldots, \varphi_{n_\xi}$, such as radial basis functions (RBFs), Fourier basis functions, or a generic dictionary [48, 49]. Thus, we have

$$f_\xi^{(\mathrm{BB})}(z) \;=\; \xi_1 \varphi_1(z) + \ldots + \xi_{n_\xi} \varphi_{n_\xi}(z), \tag{2-7}$$

for any $z \in \mathbb{R}^{n_z}$, where $z \in \mathbb{R}^{n_z}$ is the vector defined as $z = [x^\mathsf{T}, u^\mathsf{T}, d^\mathsf{T}]^\mathsf{T}$ and $n_z = n_x + n_u + n_d$ is the dimension of z. Alternatively, $f_\xi^{(\mathrm{BB})}$ can be characterized as an ℓ-layer neural network parameterized by $\xi = \mathrm{vec}(W_1, b_1, W_2, b_2, \ldots, W_\ell, b_\ell)$, where vec denotes a suitable vectorization operator, and with activation functions $\sigma_1, \ldots, \sigma_\ell$. More precisely, one can define

$$f_\xi^{(\mathrm{BB})}(z) = \sigma_\ell(W_\ell(\ldots \sigma_1(W_1 z + b_1) \ldots) + b_\ell), \tag{2-8}$$

for any $z \in \mathbb{R}^{n_z}$, where W_j, b_j, and σ_j are respectively the weight matrix, the bias vector, and the activation function in the j^{th} layer, for $j = 1, \ldots, \ell$ [50]. In addition to the parametric black-box schemes, one may employ nonparametric techniques, e.g., through Koopman operators or generic approaches [51, 52]. Specifically, given a suitably chosen matrix-valued kernel function $\Bbbk : \mathscr{M} \times \mathscr{M} \to \mathbb{R}^{n_z} \times \mathbb{R}^{n_z}$, kernel-based methods or Gaussian processes regression can be used to learn the system dynamics directly from data $\mathscr{D}$ [48]. While black-box models often provide good numerical accuracy and estimation performance, they typically lack interpretability and generalization, particularly outside the domain of the training data.

To leverage the strengths of both white-box and black-box modeling, a *two-stage hybrid approach* is proposed. The core idea is to first utilize a simplified physics-based model to capture the primary behavioral trends of the system, and subsequently, employ a black-box approach to separately learn a model for residuals from the initial stage, which effectively serves as a correction term to enhance accuracy. More precisely, within this framework, to partially embed the physics of the system directly into the model, we first consider a class of simplified physics-based models, as in (2-4), and a suitable set of black-box models, as in (2-6). Accordingly a class of *hybrid models* is defined, denoted by $\mathscr{F}^{(\mathrm{hyb})}$, as

$$\begin{aligned} \mathscr{F}^{(\mathrm{hyb})} = \{ f_{(\theta,\xi)}^{(\mathrm{hyb})} : \mathscr{M} \to \mathbb{R}^{n_x} \,| \\ f_{(\theta,\xi)}^{(\mathrm{hyb})} = f_\theta^{(\mathrm{phys})} + f_\xi^{(\mathrm{BB})}, \theta \in \Theta, \xi \in \Xi \}. \end{aligned} \tag{2-9}$$

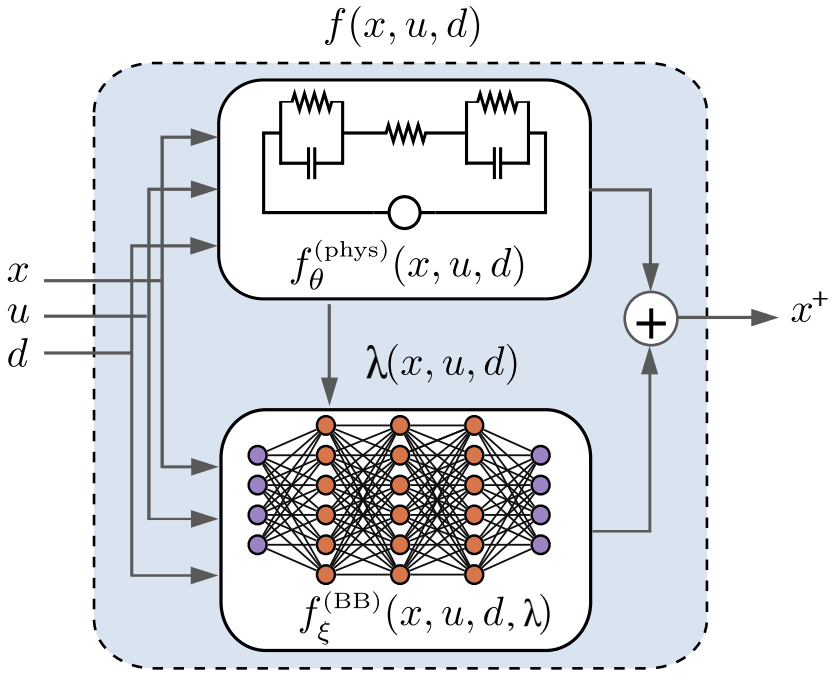

Figure 2-2: Schematic of the proposed hybrid modeling approach combining a physics-based RC circuit with neural network–based residual error estimation.

Figure 2-2 schematically illustrates the hybrid models in $\mathscr{F}^{(\mathrm{hyb})}$. Following the introduced structure for the elements of $\mathscr{F}^{(\mathrm{hyb})}$, the model identification problem reduces to estimating $\theta \in \Theta$ and $\xi \in \Xi$. Considering the flexibility of the black-box models, the parameter estimation is split into two steps, i.e., a two-stage estimation problem. First, by solving (2-4), the optimal parameter model $\theta^\star$ for the physics-based component of the model is obtained. Thus, the estimation residual terms $r_1, \dots, r_{n_\mathscr{D}}$ are defined as

$$r_k = x_{k+1} - f_{\theta^\star}^{(\mathrm{phys})}(x_k, u_k, d_k), \tag{2-10}$$

for $k = 1, \dots, n_\mathscr{D} - 1$, and form the *residual dataset* $\mathscr{R}$ as

$$\mathscr{R} = \{(x_k, u_k, d_k, r_k) \in \mathscr{M} \times \mathbb{R}^{n_x} \,|\, k = 1, \dots, n_\mathscr{D} - 1\}. \tag{2-11}$$

Subsequently, a suitable black-box modeling scheme is employed and obtain optimal parameter $\xi^\star$ using $\mathscr{R}$ and by solving an optimization problem formulated similar to (2-4). The final resulting model is

$$f_{(\theta^\star,\xi^\star)}^{(\mathrm{hyb})} = f_{\theta^\star}^{(\mathrm{phys})} + f_{\xi^\star}^{(\mathrm{BB})}. \tag{2-12}$$

Utilizing simplified physics-based models, in addition to promoting model interpretability and allowing generalization beyond the range of measurement data, leads to improving the numerical tractability of the parameter estimation problem (2-4), considering that such models are less extensively parametrized. Additionally, the two-stage structure in the proposed parameter estimation scheme is crucial, as it allows for employing flexible and representational-rich classes of black-box models, while ensuring that physics-based behavior and trends of the system are encoded in the model. More specifically, the proposed two-stage model identification

scheme, for estimating a simplified physics-based model along with a learned black-box residual, enables a balanced decomposition of the system dynamics. In the resulting model, the white-box component captures the dominant physical trends and interpretable structure, whereas the black-box residual accounts for unmodeled effects and discrepancies. Accordingly, the proposed scheme prevents the black-box model from *overfitting*, *absorbing* the entire system behavior, leaving the white-box model near zero, and resulting in an almost fully non-interpretable black-box representation, which can be the case when the parameters are jointly estimated in a single-stage approach. Thus, by separating structured knowledge from data-driven corrections, the two-stage method preserves interpretability, improves generalization, and provides a physically meaningful model. We can also regulate the bounds of the black-box models used for residual learning in various cases, e.g., by employing bounded RBF networks in (2-7), utilizing sigmoid functions for neural networks-based function representation (2-8), particularly for shallow-nets, or applying bounded kernels within nonparametric methods. Hence, one can quantify and control the deviation from the primary trend established by the physics-based component. Furthermore, through right choices of activation functions, the basis, or kernel functions, one can also establish system-theoretic guarantees and properties, namely by obtaining specific bounds, or quantifying and control- ling the deviation from the physics-based trend. Thus, the proposed scheme offers foundation for control and optimization.

Building on the proposed methodology and the preceding discussion, the next section presents the development of a hybrid model for a real-world electrolyzer using operational measurement data. To this end, the simplified physics-based component is modeled using RC circuits, while a neural network is employed to capture the residual dynamics.

2-3 Hybrid modeling for electrolyzers

This section details the application of the hybrid modeling framework from Section 2-2 to the Anion Exchange Membrane (AEM) electrolyzer. As noted in the introduction, developing a high-fidelity, first-principles model demands significant expert knowledge [26] and, crucially, requires access to specific characteristics like porosity or membrane thickness [53]. These parameters are typically proprietary and not disclosed by manufacturers of commercial units, such as the electrolyzer used in this study [54]. This makes a pure physics-based approach infeasible and makes parameter estimation from operational data computationally challenging [55]. Conversely, a purely data-driven model would sacrifice the physical interpretability that is essential for robust monitoring and control [56, 57]. To address these specific challenges, the hybrid scheme discussed in Section 2-2 and illustrated in Figure 2-2 is employed. An interpretable, simplified physics-based model in the form of a double RC equivalent circuit [44, 45] is used to capture the dominant electrochemical dynamics. A neural network-based black-box model is then combined with it to act as a targeted residual estimator, compensating for the remaining nonlinearities, unmodeled dynamics (like the thermal effects), and parameter uncertainties that the simplified RC model cannot capture.

2-3-1 Physics-based model component: an RC circuit

The equivalent circuit model encodes the key electrochemical processes within the AEM electrolyzer through physically interpretable components. Specifically, the internal resistance

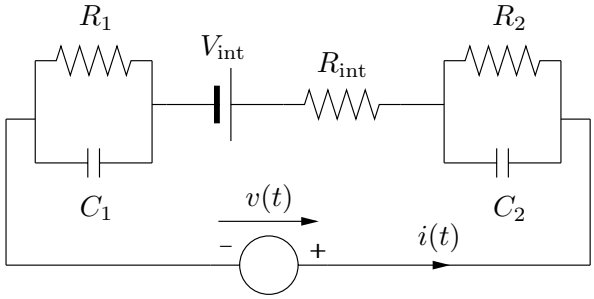

Figure 2-3: Equivalent circuit model of the AEM electrolyzer [44].

R_{int} models ohmic losses in the membrane during hydroxide ion transport, while the voltage source V_{int} represents the reversible voltage required for water splitting [44]. Moreover, each of the RC branches corresponds to a specific electrode reaction. More precisely, R_1C_1 branch represents the cathode dynamics for hydrogen evolution, as descirbed by

$$2\text{H}_2\text{O} + 2\text{e}^- \longrightarrow \text{H}_2 + 2\text{OH}^-, \tag{2-13}$$

and, R_2C_2 branch captures the anode behavior for oxygen evolution, as expressed by

$$2\text{OH}^- \longrightarrow \frac{1}{2}\text{O}_2 + \text{H}_2\text{O} + 2\text{e}^-. \tag{2-14}$$

The capacitances represent the electrical double-layer at each electrode interface, accounting for the charge separation that governs transient response during current variations [58]. The distinct time constants of the two RC branches reflect the different reaction kinetics, with typically faster hydrogen evolution at the cathode compared to oxygen evolution at the anode [59].

2-3-2 Parameter estimation for the equivalent RC circuit

For the dynamics of the double-branch RC circuit shown in Figure 2-3, we have

$$\frac{\text{d}}{\text{d}t}\begin{bmatrix} V_{\text{RC1}}(t) \\ V_{\text{RC2}}(t) \end{bmatrix} = \begin{bmatrix} -\frac{1}{\tau_1} & 0 \\ 0 & -\frac{1}{\tau_2} \end{bmatrix}\begin{bmatrix} V_{\text{RC1}}(t) \\ V_{\text{RC2}}(t) \end{bmatrix} + \begin{bmatrix} \frac{R_1}{\tau_1} \\ \frac{R_2}{\tau_2} \end{bmatrix} i(t), \tag{2-15}$$

where V_{RC1} and V_{RC2} are the voltages across each RC branch, τ_1 and τ_2 are the time constants, R_1 and R_2 are their resistances, and $i(t)$ is the current driving the electrolyzer. Assume that $i(t)$ has a piece-wise constant structure with long-enough time-intervals $[s^j, s^{j+1})$. Let I_j denote the value of $i(t)$ during the j^{th} time-interval. Accordingly, one can approximate the electrolyzer voltage at $t \in [s^j, s^{j+1})$ as

$$v(t) = (V_0^{(j)} - V_{\text{ss}}^{(j)})\text{e}^{-\frac{1}{\tau}(t-s^j)} + V_{\text{ss}}^{(j)}, \tag{2-16}$$

where $V_0^{(j)}$ and $V_{\text{ss}}^{(j)}$ are the initial and steady-state voltages. Furthermore, for any $t \in [s^j, s^{j+1})$, the following is known:

$$V_{\text{RC1}}(t) = (V_{\text{RC1,0}}^{(j)} - V_{\text{RC1,ss}}^{(j)})\text{e}^{-\frac{1}{\tau_1}(t-s^j)} + V_{\text{RC1,ss}}^{(j)}, \tag{2-17}$$

$$V_{\mathrm{RC2}}(t) = (V^{(j)}_{\mathrm{RC2},0} - V^{(j)}_{\mathrm{RC2,ss}})\mathrm{e}^{-\frac{1}{\tau_2}(t-s^j)} + V^{(j)}_{\mathrm{RC2,ss}}. \tag{2-18}$$

Moreover, for any t, it is also known that

$$v(t) = V_{\mathrm{RC1}}(t) + V_{\mathrm{int}} + i(t)R_{\mathrm{int}} + V_{\mathrm{RC2}}(t), \tag{2-19}$$

where V_{int} denotes the internal activation voltage. The time constants, resistances, and capacitances are jointly related as

$$\tau_1 = R_1 C_1, \tag{2-20}$$
$$\tau_2 = R_2 C_2, \tag{2-21}$$
$$R_{\mathrm{tot}} = R_{\mathrm{int}} + R_1 + R_2, \tag{2-22}$$

Note that (2-22) implies $R_{\mathrm{int}} = R_{\mathrm{tot}} - R_1 - R_2$, and therefore, R_{int} can be determined by estimating R_{tot}, R_1, and R_2, provided that $R_{\mathrm{tot}} - R_1 - R_2 \geq 0$. Accordingly, let $\theta = [R_1, R_2, R_{\mathrm{tot}}, C_1, C_2, V_{\mathrm{int}}, \tau_1, \tau_2]^{\mathsf{T}}$ be the vector of parameters. Following the above derivations, the function $V_{\mathrm{sim}}(\theta, i(t))$ is defined as the mathematical relation that approximately *simulates* the ground-truth voltage $v(t)$. Considering discrete-time measurements (i_k, v_k), the parameter estimation problem is formulated as

$$\begin{aligned} \min_{\theta \in \mathbb{R}^{n_\theta}_{\geq 0}} \quad & \sum_{k=1}^{n_k} \left(V_{\mathrm{sim}}(\theta, i_k) - v_k\right)^2 \\ \text{s.t.} \quad & R_{\mathrm{tot}} - R_1 - R_2 \geq 0, \\ & \text{(2-20) and (2-21).} \end{aligned} \tag{2-23}$$

Since the estimation problem in (2-23) is non-convex, its solution is prone to being trapped in local spurious minima. Consequently, to solve (2-23), an alternative tractable multi-step approach needs to be employed, which is discussed below. First, linear regression in the log-scale domain is used on transient data to obtain τ_1 and τ_2. A similar regression on the polarization curve yields the total resistance R_{tot} and the internal activation voltage V_{int} [44]. With these parameters known, the individual branch resistances (R_1, R_2) can be found by solving a simplified version of (2-23) formulated as

$$\begin{aligned} \min_{R_1, R_2 \in \mathbb{R}_{\geq 0}} \quad & \sum_{k=1}^{n_k} \left(V_{\mathrm{sim}}(R_1, R_2, i_k) - v_k\right)^2 \\ \text{s.t.} \quad & R_{\mathrm{tot}} - R_1 - R_2 \geq 0. \end{aligned} \tag{2-24}$$

To provide a clear and reproducible method, the entire parameter estimation procedure, adapted from the methodology presented by Guilbert and Vitale [44], which combines these linear regression and numerical optimization steps, is summarized in Algorithm 1.

Algorithm 1 systematically breaks down the estimation process. **Part 1** determines the static parameters by fitting the polarization curve. **Part 2** isolates the dynamic behavior by analyzing the transient voltage response; it segments the data to distinguish between fast and slow dynamics, applying a logarithmic transformation to linearize the exponential decay and extract the time constants τ_1 and τ_2. **Part 3** uses these values to solve the constrained optimization problem from (2-24), determining the individual resistances R_1 and R_2. Finally, **Part 4** computes the remaining parameters, ensuring that the physical relationships in (2-20)-(2-22) are satisfied. Considering the low dimension of the optimization variable in this

Algorithm 1 Parameter Estimation for the Double-Branch RC Model.

Input: Polarization dataset $\mathcal{D}_{\text{pol}} = \{(I_k, V_k)\}_{k=1}^{N_{\text{pol}}}$; Full transient dataset $\mathcal{D}_{\text{trans}} = \{(t_k, I_k, V_k)\}_{k=1}^{N_{\text{trans}}}$ after a current step.
Output: The complete parameter set $\theta = \{R_1, C_1, R_2, C_2, R_{\text{int}}, V_{\text{int}}\}$.
▷ ***Part 1: Identify static parameters from polarization data*** ◁
Perform linear regression on $\mathcal{D}_{\text{pol}}$ to fit $V_k = I_k \cdot R_{\text{tot}} + V_{\text{int}}$.
Obtain total resistance R_{tot} (slope) and reversible voltage V_{int} (intercept).
▷ ***Part 2: Identify time constants from transient data*** ◁
Segment $\mathcal{D}_{\text{trans}}$ into a fast transient subset $\mathcal{D}_{\text{fast}}$ and a slow transient subset $\mathcal{D}_{\text{slow}}$.
function FINDTAU($\mathcal{D}_{\text{subset}}$)
    Determine the final steady-state voltage V_∞ for the subset.
    Create a linearized dataset $\mathcal{D}_{\text{lin}} = \{(t_k, \ln(V_\infty - V_k))\}$ for $(t_k, V_k) \in \mathcal{D}_{\text{subset}}$.
    Perform linear regression on $\mathcal{D}_{\text{lin}}$ to find the slope m.
    return $-1/m$.
$\tau_1 \leftarrow$ FindTau($\mathcal{D}_{\text{fast}}$).
$\tau_2 \leftarrow$ FindTau($\mathcal{D}_{\text{slow}}$).
▷ ***Part 3: Identify dynamic resistances via optimization*** ◁
Define an objective function $J(R_1, R_2)$ which calculates the MSE between the measured voltage from $\mathcal{D}_{\text{trans}}$ and the voltage from a simulated RC model, $V_{\text{sim}}(t; R_1, R_2)$.
    ▷ *The simulation uses known V_{int}, τ_1, τ_2 and the constraint $R_{int} = R_{tot} - R_1 - R_2$.*
Solve the constrained optimization problem:
$(R_1^*, R_2^*) \leftarrow \arg\min_{R_1,R_2} J(R_1, R_2)$
subject to: $R_1 > 0, R_2 > 0, R_1 + R_2 < R_{\text{tot}}$.
▷ ***Part 4: Calculate final parameters*** ◁
$R_{\text{int}}^* \leftarrow R_{\text{tot}} - R_1^* - R_2^*$.
$C_1^* \leftarrow \tau_1 / R_1^*$.
$C_2^* \leftarrow \tau_2 / R_2^*$.
return $\{R_1^*, C_1^*, R_2^*, C_2^*, R_{\text{int}}^*, V_{\text{int}}\}$.

problem, it is computationally tractable despite its nonconvex nature. Standard approaches, such as sequential quadratic programming, can efficiently find a solution that satisfies both the objective function and the physical constraints.

It should be noted that the parameter identification for ECMs, namely the low-order RC models [27, 28, 30–35], regardless of the branch order, fundamentally reduces to a robust linear constrained optimization problem, making the final V − I emulation accuracy primarily dependent on the physical fidelity of the RC structure and the dynamic content of the training data.

2-3-3 Physics based thermal dynamics

The thermal behavior of the electrolyzer is governed by a balance of heat generation, passive heat dissipation to the environment, and active cooling from its internal closed-loop system. Following [60], this relationship can be represented by the differential equation

$$\frac{dT}{dt} = K_1\dot{Q}_{\text{gen}} - K_2\dot{Q}_{\text{loss}} - K_3\dot{Q}_{\text{cool}} \tag{2-25}$$

where $\dot{Q}_{\text{gen}}$ is the heat generated by the stack's inefficiency (proportional to the input power), $\dot{Q}_{\text{loss}}$ represents passive heat loss, and $\dot{Q}_{\text{cool}}$ is the heat removed by the cooling system.

For control-oriented purposes, this physical relationship can be effectively captured by a discrete-time linear model. The model predicts the stack temperature at the next time step (T_{k+1}) based on the current temperature (T_k) and the electrical power input (P_k^{el}):

$$T_{k+1} = g_{\text{temp}}T_k + g_{\text{gain}}P_k^{\text{el}} + g_{\text{base}} \tag{2-26}$$

Here, the parameters g_{temp}, g_{gain}, and g_{base} are identified from operational data and represent the combined effects of thermal inertia, heat generation, and cooling. This linear model provides a computationally efficient yet accurate method for dynamically estimating the stack temperature [60]. When training on 80 % the data, the trend in temperature is very clearly visible. While the closed loop control system in the EL could be modelled as a hybrid model, this simple linear approximation is good enough at this time. The results are shown in Figure 2-4.

2-3-4 Neural network for residual estimation: capturing the temperature Influence

The operational temperature significantly influences the performance of the AEM electrolyzer by affecting electrochemical kinetics, the Nernst potential, and the ionic conductivity of the membrane [61]. While our physics-based RC model provides an interpretable foundation for the electrolyzer's primary dynamics, its fixed-parameter nature makes it incapable of accounting for these thermal effects [62].

To address this structured deficiency, a neural network (NN) is employed to estimate the residual error, as outlined in the proposed hybrid framework. TheNN is specifically designed to learn the complex, nonlinear relationship between the operational inputs (current and temperature) and the error signal from the RC model. This synergy is precisely how the hybrid approach achieves a high-fidelity representation:

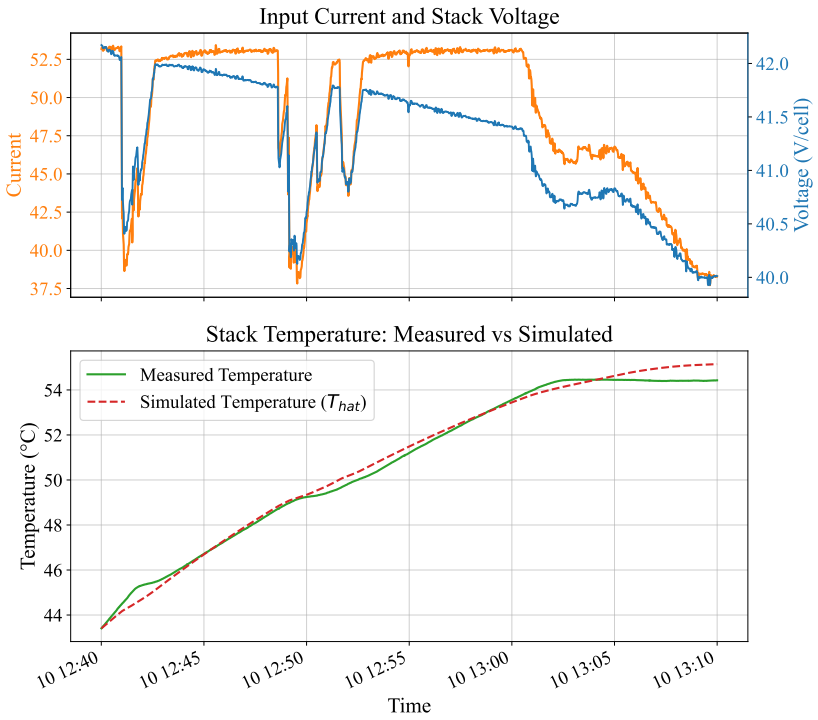

Figure 2-4: The temperature model derived based on Equation (2-26).

The RC circuit preserves the model's physical interpretability, capturing the dominant, well-understood electrochemical behavior. The neural network acts as a targeted correction factor, compensating only for the unmodeled, temperature-dependent dynamics that the simplified physics-based component cannot capture.

By combining the physically-grounded structure of the RC model with the adaptive, data-driven accuracy of the NN, the resulting hybrid model can more accurately represent the true system behavior under real-world operating conditions. Further details on the implementation and validation of this methodology are presented in the next section through a case study of an electrolyzer unit installed at The Green Village research site.

2-4 Hybrid modeling results for an AEM electrolyzer

2-4-1 Electrolyzer specifications and operational conditions data

The operational data was collected from the Enapter EL 2.1 unit (specifications in Table II) within the 24/7 energy hub, sampled at 1 Hz, i.e., $\Delta = 1\,\text{s}$. Analysis of the operational data confirmed that the internal control system maintains a tight cathode pressure (mean 20.2 bar, 1.02% relative standard deviation) and near-ambient anode pressure (mean 1.74 bar). As these pressures are stable and well-regulated, their dynamic influence on voltage is minimal. With the goal of real-time control, e.g., implementing a suitable MPC strategy, the main focus can be only on obtaining the dynamic model depending primarily on the externally-driven variables, namely the fluctuating current density and the resulting slower thermal dynamics.

Table 2-1: Key parameters of the Enapter EL 2.1 AEM electrolyzer.

Parameter	Value
Electrical	
Supply Voltage	200–240 V AC
Nominal Power Consumption	2.4 kW
Peak Power Consumption	3.0 kW
Efficiency	4.8 kWh/Nm^3 H_2
Operating Conditions	
Operating Temperature	5–45°C (Ambient)
Electrolyte Temperature	55°C (Nominal)
Production Rate Range	60–100% (of Nominal)
Nominal H_2 Production	0.5 Nm^3/hr
Output Pressure	Up to 35 bar
Material and Inputs	
Water Consumption	$\sim$ 400 mL/hr
Max Water Conductivity	< 20 μS/cm
Process Liquid	1% KOH solution
Physical	
Dimensions (W $\times$ D $\times$ H)	482 $\times$ 634 $\times$ 307 mm
Weight	55 kg

2-4-2 Signal processing and time constant determination

Since the signal is relatively noisy, it is first preprocessed using a Butterworth filter, with a cutoff frequency $\omega = 0.05\,\text{rad/sample}$. Given that the sampling frequency is $F_s = 1\,\text{Hz}$, the employed cutoff frequency corresponds to 0.05 Hz. This choice effectively attenuates high-frequency measurement noise while preserving the integrity of the underlying dynamics crucial for accurate model identification. Moreover, by solving the regression problem discussed in Section 2-3-2, the time constants can be estimated for both the fast transient and the slower stabilizing transient as $\tau_1 = 2.6\,\text{s}$ and $\tau_2 = 123.8\,\text{s}$, respectively.

2-4-3 Resistance and activation voltage determination

The relationships between time constants, resistances, and capacitances are explained in equations (2-20), (2-21), (2-22). Following the discussion in Section 2-3-2, to derive the total resistance R_{tot} and the internal activation voltage V_{int}, one can employ linear regression on the polarization curve of the stack. Therefore, from Figure 2-5, the corresponding values for the system can be obtained as

$$R_{\text{tot}} = 1.31\ \Omega \tag{2-27}$$

$$V_{\text{int}} = 1.5683\ \text{V}. \tag{2-28}$$

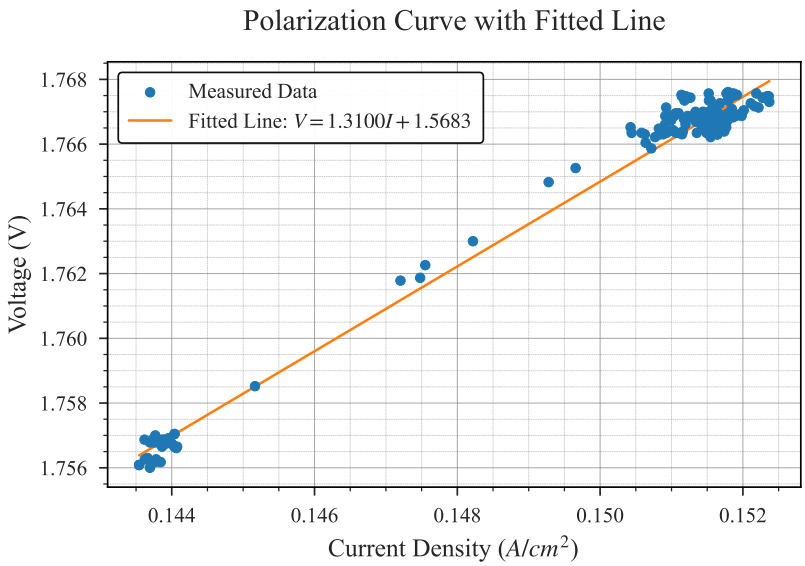

Figure 2-5: Polarization curve fitting applied to the AEM electrolyzer stack.

2-4-4 Individual branch resistance calculation

Following the robust, multi-step estimation procedure from Algorithm 1, the constrained optimization (2-24) was solved to find the individual branch resistances. This sequential identification yields $R_2 = 0.197 \times 10^{-3}\Omega$ and $R_1 = 9.32 \times 10^{-3}\Omega$. This approach ensures the parameters retain their physical interpretability, a key advantage over monolithic or black-box methods.

It is important to note that this sequential estimation procedure, detailed in Algorithm 1, is key to the robustness of the identification. By first determining static parameters $(R_{\text{tot}}, V_{\text{int}})$ and time constants (τ_1, τ_2) from distinct data features, the final optimization for R_1 and R_2 is low-dimensional and well-constrained. This avoids the numerical instability and local minima pitfalls of a large-scale, all-in-one optimization, resulting in a more robust physical model.

2-4-5 Model enhancement via neural networks

The simulation results, as shown in Figure 2-6, demonstrate that the RC model successfully captures the fundamental dynamic behavior of the electrolyzer, including the initial fast transient and the subsequent slower voltage evolution. However, a persistent and structured error between the simulated and measured voltage is clearly visible. However, a persistent and structured error is visible in Figure 2-6. Cross-correlation analysis confirms this residual is strongly and inversely correlated with temperature, validating the need for a targeted, temperature-dependent corrector. Comparison with different state-of-the-art modeling methodologies is presented in the next section.

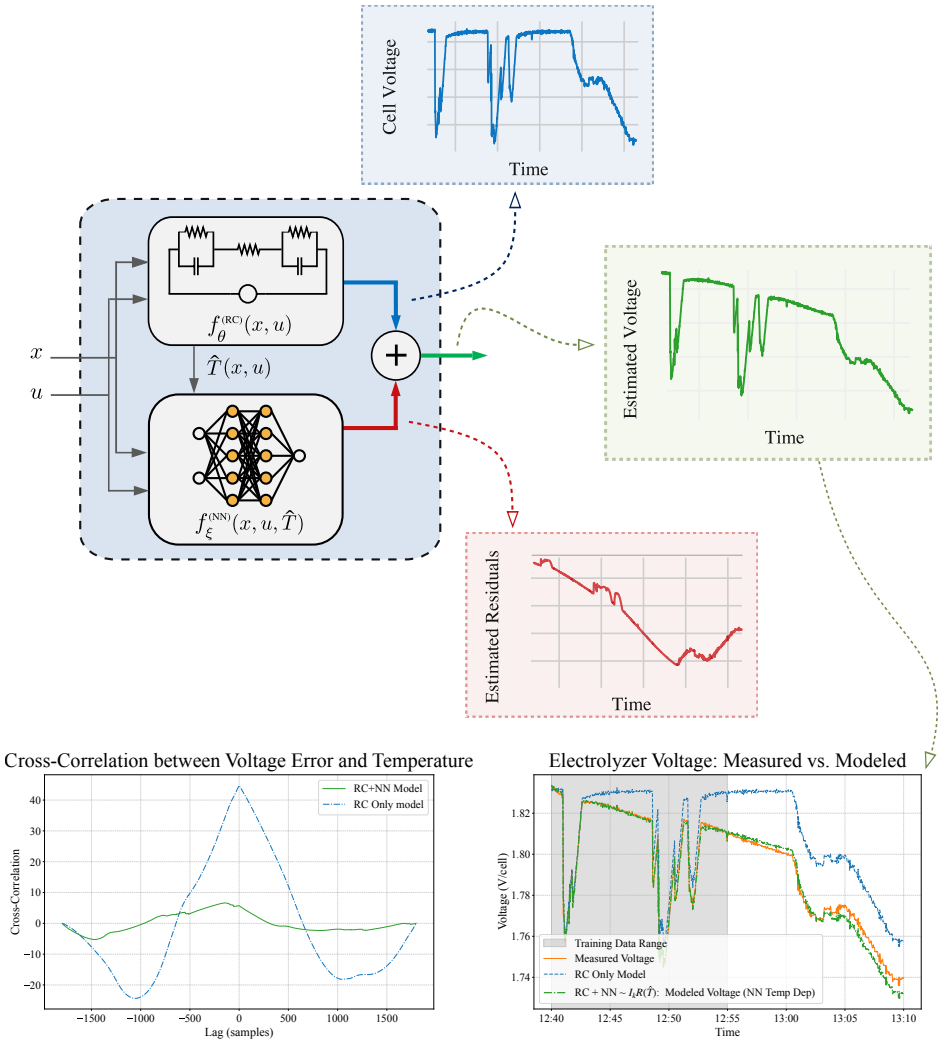

Figure 2-6: Schematic of the proposed hybrid model architecture; comparison of the RC model and the hybrid model for the error signal; and comparison of their voltages against the measured voltage.

2-4-6 Results and Comparison

To validate the hybrid $\mathrm{RC} + \mathrm{NN}(I_k, \hat{T}_k)$ model, its performance was rigorously benchmarked against the baseline RC model and several other advanced modeling architectures. Performance was quantified using the model fit percentage.

Let N_T be the total number of data samples in the validation set, y_k represent the measured voltage at time step k, and $\hat{y}_k$ be the voltage predicted by the model at the same time step. Define the mean of the measured signal, denoted by $\bar{y}$, as

$$\bar{y} = \frac{1}{N_T} \sum_{k=1}^{N_T} y_k \tag{2-29}$$

The model fit percentage is a normalized metric that compares the model's prediction error to

the total variance of the data. It is calculated using as

$$\text{fit } (\%) = 100 \times \left(1 - \frac{\sqrt{\sum_{k=1}^{N_T} (y_k - \hat{y}_k)^2}}{\sqrt{\sum_{k=1}^{N_T} (y_k - \bar{y})^2}} \right) \tag{2-30}$$

The results from Table 2-2 are conclusive. The baseline RC model alone achieves a poor fit of only 26.00 %, confirming its structured error makes it inadequate for high-fidelity applications. While introducing a simple linear temperature correction (RC + $I_k(\alpha + \beta\Delta\hat{T}_k)$) dramatically improves the fit to 79.76%, reinforcing that temperature is the primary source of error, the hybrid model achieves the highest performance with a 92.03 % fit. This represents a 66% absolute improvement over the baseline and, crucially, also outperforms other data-driven models, including the pure black-box DNN(I_k) (52.04% fit) and the PINN(I_k) (74.91% fit).

This superior performance is a direct result of the hybrid design, which balances all key objectives. The high accuracy is achieved with a computationally tractable model; the shallow network structure (Table 2-2) ensures the model can be evaluated rapidly, making it ideal for real-time, optimization-based control like MPC. This deliberate low parametrization is a key feature of the robust learning approach, as the network is less prone to overfitting and is focused only on learning the structured, temperature-dependent residual. Finally, the two-stage estimation strategy employs the RC circuit as an interpretable backbone, preserving its physical meaning, an advantage lost in the pure DNN andPINN approaches. The results confirm that temperature fluctuations cause significant systematic errors in the RC model, validating the necessity of a targeted hybrid correction.

Table 2-2: Performance Comparison of Modeling Approaches and Architectures

Model ID	**Architecture**	**Model Fit (%)**
RC model:	Double RC ECM	26.00
RC + $I_k(\alpha + \beta\Delta\hat{T}_k)$	RC and $R_{\text{res}}(\hat{T})$	79.76
RC + $I_k \cdot \text{NN}(\hat{T}_k)$	[10] + ReLU	87.28
here: $\text{NN}(\hat{T}_k) \approx \Delta R_{\text{int}}(\hat{T}_k)$	[100] + ReLU	83.65
	[130] + tanh	91.09
RC + $\text{NN}(I_k, \hat{T}_k)$	[3]	74.27
	[10]	92.03
	[10-10]	80.93
DNN(I_k)	[64, 64, 32] + tanh	52.04
PINN(I_k)	[30, 30, 30] + tanh	74.91

2-4-7 Discussion

The results validate the proposed hybrid methodology as a robust, control-oriented modeling framework. The baseline RC model's 26.00% fit confirms it is inadequate for high-fidelity applications, as it fails to model the large, systematic, temperature-dependent errors. The

proposed hybrid model, by successfully training a lightweight neural network on this residual, achieved a 92.03% fit. This result's significance is highlighted by the benchmark in Table 2-2, as the model decisively outperforms both a pure black-box DNN (52.04% fit) and a traditional physics-in-loss PINN (74.91% fit).

This two-stage estimation strategy is a deliberate choice over simultaneous optimization [38], as it locks in the interpretable RC parameters first, preventing theNN from overpowering the physics. This methodology achieves the best of both worlds for real-time control: the computational efficiency and physical interpretability of the RC model, combined with the high accuracy of a data-driven corrector, making it a compelling candidate for advanced applications like MPC. While this 1 Hz model is highly accurate, the small remaining error points to future work, such as online parameter estimation to account for long-term stack degradation or modeling other unmeasured physical variables.

The findings demonstrate that the combination of physics-based and data-driven models increases the models ability to represent electrolyzer stack dynamics, which makes the hybrid model a compelling candidate for use in real-time estimation, diagnostic monitoring, and control applications, especially in systems where temperature effects and operational variability cannot be ignored.

2-5 Conclusion and future directions

In this paper, a novel hybrid model is successfully developed and validated for a commercial anion exchange membrane electrolyzer using real-world operational data. It was demonstrated that combining a double-branch Resistor-Capacitor (RC) circuit with a simple neural network for residual estimation provides a high-fidelity representation of the electrolyzers dynamic behavior. The key achievement of this work is the significant improvement in predictive accuracy; the hybrid model increased the model fit by 66 % compared to the conventional RC model alone. Our analysis revealed that this improvement is primarily due to the neural network's ability to effectively learn and compensate for systematic, temperature-dependent dynamics that the fixed-parameter RC model could not capture. The proposed methodology achieved a crucial balance between the physical interpretability of the RC circuit and the accuracy of a data-driven approach, without the computational burden of complex multiphysics models. Future work will focus on integrating this high-fidelity model into a MPC framework for real-time optimization of the electrolyzer within the HES system. Subsequent research will investigate extending the hybrid approach to capture faster, millisecond-scale dynamics, which requires higher-frequency data acquisition and is relevant for grid stability studies. Additionally, adaptive control strategies, such as switching models based on operational regimes, will be explored to balance computational efficiency and predictive accuracy.

Chapter 3

Data-Driven Modeling of LV Distribution Grids: A Case Study in Bonaire

3-1 Introduction

The global energy landscape is undergoing a profound transformation, driven by the imperative to mitigate climate change by transitioning away from fossil fuels [1]. This shift is characterized by the increasing reliance on Renewable Energy Sources (RES) such as solar and wind. However, the inherent intermittency of these sources poses a significant challenge to the stability and reliability of electrical grids, creating a potential mismatch between variable power generation and fluctuating consumer demand [2]. For Distribution Service Operators (DSOs), this rapid adoption of decentralized energy generation introduces a host of new operational challenges, including bidirectional power delivery, phase imbalance on different phases, and unexpected transformer loading [63].

To address this intermittency, Energy Storage Systems (ESS) have emerged as a critical enabling technology, allowing for the large-scale integration of renewables by balancing energy supply and demand [3]. Hybrid Energy Storage (HES) systems, which combine different storage types (e.g., fast-response batteries with long-duration hydrogen), are particularly promising for enhancing grid stability [4]. However, the integration of these advanced, decentralized assets introduces significant new complexities in control and grid interaction that must be effectively managed to prevent instability [9].

This is where the importance of accurate network models becomes paramount. Effective planning, placement, and control of HES systems and other Distributed Energy Resources (DER) are entirely dependent on a reliable and accurate model of the distribution grid. These models function as "digital twins", which are essential for simulating power flows and the grid's response to storage operations. DSOs rely on these models to perform advanced simulations and analyses, such as assessing the hosting capacity for new DERs [7], optimizing battery

dispatch to prevent grid congestion [9], and evaluating and prioritizing crucial investment and operational decisions [10].

Effectively addressing these operational and planning challenges requires a deep understanding of distribution grid characteristics. However, research comprehensively detailing these characteristics, particularly for dynamic, DER-rich environments, remains relatively sparse, but there are some examples in the EU and Brazil: [64], [65], and [7]. However, grid characteristics often heavily depend on their geographical location. This is closely due to user behavior, weather related characteristics and a different customer base. In the Caribbean region, there is not a lot of data available on these distribution loading statistics.

This paper aims to combine existing techniques into a data driven modeling pipeline which can be used to simulate distribution grids. In section 3-2, the theoretical background is given for the different aspects of generating power models. This background is used to generate methodology in Section 3-3, which is then applied as a case study in Section 3-4. Section 3-5 contains the results for the numerical experiments, where the simulation results are validated against real world measurements.

3-2 Theoretical Background & Literature Review

All grid models, which are central tools for managing and planning electricity systems, require accurate input data encompassing the grid's structure and its electrical properties. The first critical requirement for developing a coherent grid model is the reconstruction of the network topology [66], which involves defining the network of nodes and interconnections (power lines) [67]. Recent approaches to addressing data availability issues focus on deriving this topology transparently using open data sources like OpenStreetMap (OSM), employing techniques such as attribute-based or geometry-based abstraction. The second fundamental requirement, electrical component modeling, involves defining the physical and electric parameters of these network elements [67].

3-2-1 Network Topology Identification

Topology derivation approaches, often referred to as Topological Path Identification (TPI) or Topology Identification (TI), are broadly characterized by their input data. They fall into two main categories: methods relying only on static data (such as GIS maps) and those leveraging dynamic measurements from an Advanced Metering Infrastructure (AMI) [10].

The first category, which is most relevant to limited-data environments, must address the typical inaccuracies and incompleteness of static GIS data. To solve this, optimization techniques have been introduced, such as formulating the TPI problem as an Integer Linear Programming (ILP) algorithm [10]. Conceptually, these methods assign a 'cost' c_{ij} (e.g., the geographical length) to each potential line segment between nodes i and j. A binary decision variable, x_{ij}, represents the inclusion of that segment in the final topology. The core objective is to find a radial (tree) network that minimizes the total cost, as shown in Equation (3-1), while adhering to known constraints, such as customer-to-transformer connections.

$$\min \sum_{(i,j)\in E} c_{ij} \cdot x_{ij} \tag{3-1}$$

In cases where the set of buses served by each substation is unknown, a two-step approach may be required: first, performing bus clustering to partition the network, which can be framed as a non-convex optimization problem solved using techniques like the Alternating Direction Method of Multipliers (ADMM); and second, applying a TPI algorithm within those partitioned areas [10].

In contrast, the second major category of methods relies on dynamic AMI data. These approaches often use rigorous optimization algorithms, such as Mixed-Integer Linear Programming (MILP) or Mixed-Integer Quadratic Programming (MIQP) [68], to estimate the topology from smart meter or line sensor recordings [69]. To tackle the uncertainty caused by DERs, hybrid frameworks have been developed that combine model banks with high-dimensional analytics, using tools like Auto-regression (AR) and Random Matrix Theory (RMT). An alternative "active" approach is power grid probing, where smart inverters intentionally perturb power injections and the resulting voltage deviations are analyzed using graph algorithms to recover the topology [66, 70].

Furthermore, topology can be learned using probabilistic graphical models, where dependencies among voltage measurements are described. This allows the topology to be found by identifying the maximum weight spanning tree from mutual information-based correlations. For LV grids with unmonitored "latent" nodes, a specialized Latent Tree Model (LTM) [71] can be used. This Bayesian network employs a search algorithm guided by the Bayesian Information Criterion (BIC) and enhanced by the Expectation-Maximization (EM) algorithm to estimate the full topology using only end-user smart meter data.

3-2-2 Power Grid Component Modeling

Since the modeling of electrical power systems has been around for some time, different approaches are available such as copper plate models, where the topology can be neglected. Other options include (linear) DC models: more sophisticated models that define a network of nodes and interconnections (power lines). It uses Kirchhoff's law to determine active power flows, which depend on the resistance (R) and maximal capacity of the power lines. Even more accurate are AC power flow model, that also take reactance into account [67].

The development of electrical component models of the power grid often relies on physical analogies and the application of foundational electrical engineering concepts [72]. For distribution grids, the analysis frequently uses linearized models where bus voltages are expressed as linear functions of active (P) and reactive (Q) power injections. This results in models like the Linearized Distribution Flow (LDF) model, which defines nodal voltage magnitudes using matrices (R, X) that are derived from the grid's weighted reduced Laplacian matrices [66, 73].

However, for more accurate results, particularly when assessing reliability or vulnerability, modern approaches integrate physical and electrical parameters into the topological models. These enriched models use parameters such as impedance for transmission lines and power limits for substations. When performing resilience analysis, models that incorporate physical parameters, such as simplified direct current (DC) power flow models [72], provide a more realistic view of how failures disrupt and spread through the system compared to purely topological analyses. Furthermore, within high-complexity systems like integrated circuit power/ground (P/G) networks, modeling can be sped up by using equivalent circuit modeling to construct simplified representations of many series resistor circuits, leveraging the regularity

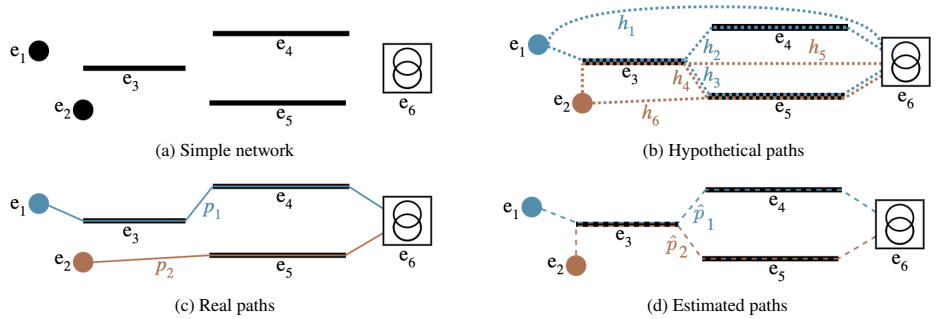

Figure 3-1: TPI technique and figure adapted from [10]

(e.g., constant width) often found in these networks [74]. Since the accuracy of these component models determines the credibility of simulation results, obtaining precise circuit parameters, whether through experiments or theoretical calculation based on the component's physical characteristics, is essential.

3-3 Methodology: A Pipeline for Limited-Data Grid Modeling

3-3-1 Stage 1: Network Topology Reconstruction

A correct network topology is important to create a representative network. Accurately identifying how customers are electrically connected (their topological paths), despite being often not completely clear, is essential for many modern power system applications. This knowledge forms the foundation for creating a network digital twin.

To determine the most probable electrical paths from the available static data, this work adopts the methodology proposed by Vassallo et al. [10]. The optimization algorithm begins with data preparation that uses only static data, such as the GIS coordinates of network elements, their types (e.g., customer, line, junction), and the known feeder terminal junction each customer is connected to. This raw data is used to generate a set of "hypothetical paths". For large networks where listing all possible paths is "impractical", the method employs an A* algorithm. This heuristic search algorithm efficiently approximates the set of viable hypothetical paths by exploring connections for each customer to their designated terminal junction, respecting constraints like maximum distances. The resulting list of paths is then converted into binary matrices (H_C, H_R, H_T) that serve as the main parameters for the ILP problem. This approach is shown in Figure 3-1.

Paths are representations of the route that electricity will follow from the transformer terminal to the customer. There are three different paths considered:

The complete explanation of the algorithm can be found in [10], but for the completeness we will consider a short overview; the algorithm goes through the following three key stages:

1. Data Preperation: the raw data is transformed into three sets: customers $\mathcal{C}$, terminal elements $\mathcal{T}$, and other network elements such as lines $\mathcal{R}$.

Term	Description
Hypothetical paths ($\mathcal{H}$)	The set of all potential sequences of network elements through which electricity might flow to supply a customer.
Real paths ($\mathcal{P}$)	The set of actual sequences of network elements connecting a customer to a specific source (e.g., an MV/LV transformer). This is a subset of the hypothetical paths, i.e., $\mathcal{P} \subseteq \mathcal{H}$.
Estimated paths ($\hat{\mathcal{P}}$)	The set of paths determined through estimation or modeling techniques, intended to approximate the real paths $\mathcal{P}$. This set is also considered a subset of the hypothetical paths, i.e., $\hat{\mathcal{P}} \subseteq \mathcal{H}$.

Table 3-1: Definitions of different path sets in the network, defined by [10]

2. Based on the structured element data and connectivity rules (e.g., potential connections based on distance, radial network structure), a comprehensive set of all plausible paths $\mathcal{H}$ from each customer to their designated feeder terminal junction is generated, using the A* algorithm.
3. The problem of selecting the best estimate of the real paths ($\hat{\mathcal{P}}$) from the compatible hypothetical paths ($\mathcal{H}^{\mathcal{I}'}$) is formulated as an ILP optimization problem.

The goal is to maximize the number of identified customer paths, with a penalty for the number of elements assigned to terminals. This results in objective function (3-2).

$$\max_{\hat{P}, T_R} \quad \omega \cdot \left(\sum_{k=1}^{|\mathcal{H}^{I'}|} \hat{P}_{(k)} \right) - \sum_{m=1}^{|\mathcal{T}|} \sum_{n=1}^{|\mathcal{R}|} T_{R(m,n)} \tag{3-2}$$

where ω is a weight balancing path identification and element assignment.

The key constraints ensure the feasibility and logical consistency of the solution:

- **Path Validity:** All elements within a selected path must belong to the same terminal (feeder junction).
- **Unique Customer Path:** Each customer can be associated with at most one estimated path.
- **Unique Element Assignment:** Each intermediate element ($\mathcal{R}$) can be assigned to at most one terminal.

3-3-2 Stage 2: Grid Asset Parameterization

The topological graph derived in the previous Section provides the "skeleton" of the distribution grid. To convert this static map into a dynamic grid model suitable for simulation, a precise electrical model must be assigned to each component. This process is the core of the modeling pipeline.

The objective is to translate the physical properties of assets (lines, transformers, loads) into a system of equations that can be solved for unknown variables, namely the voltage magnitude and angle at each node. This entire framework is implemented using the `power-grid-model` (PGM) library [11], which provides a robust structure for steady-state analysis. Power Grid Model is a high-performance Python/C++ library for steady-state distribution power system analysis, developed as an open source project hosted by the LF Energy foundation. The main contributors are DSO's, universities, research institutes and commercial parties [75].

Power System Analysis

The foundation of steady-state power system analysis is the linear relationship between the vector of nodal current injections (I_{bus}) and the vector of nodal voltages (V_{bus}) through the nodal admittance matrix, Y_{bus}.

$$I_{bus} = Y_{bus} \cdot V_{bus} \tag{3-3}$$

Where:

- V_{bus} **(The Unknowns):** A complex vector $[V_1, V_2, ..., V_n]^T$, where each element $V_i = |V_i|\angle\delta_i$ is the complex voltage at a node (bus). These are the primary variables we aim to solve for.
- I_{bus} **(The Inputs):** A complex vector $[I_1, I_2, ..., I_n]^T$, where each I_i is the net complex current injected into node i. This is determined by the loads and generators.
- Y_{bus} **(The System Model):** An $n \times n$ matrix representing the grid's physical connections and passive properties. This is the matrix we must **build** by parameterizing each asset.

The Y_{bus} matrix is assembled by applying Kirchhoff's Current Law. Each component's admittance ($y = 1/z$) is "stamped" into the matrix according to two rules:

1. **Diagonal Elements (Y_{ii}):** The "self-admittance" of node i. It is the **sum** of all admittances of components connected *to* node i.

$$Y_{ii} = \sum_{j \neq i} y_{ij} + y_{i,\text{shunt}} \tag{3-4}$$

2. **Off-Diagonal Elements (Y_{ij}):** The "mutual admittance" between nodes i and j. It is the **negative** of the sum of all admittances connected *between* nodes i and j.

$$Y_{ij} = -\sum y_{ij} \tag{3-5}$$

Transformer

A transformer is a branch which connects two nodes with different voltage levels. The transformer is represented as a π model, consisting out of two branches:

- A series branch representing the impedance (Z_{series}) due to the windings' resistance and leakage reactance. This accounts for the voltage drop and losses when current flows through the transformer.
- A shunt branch representing the admittance (Y_{shunt}) across the terminals. This accounts for the magnetization characteristics (reactance) and core losses (resistance).

Both these branches are represented as two sets of equations, which can be found in the documentation for the transformer model in [11].

Junction

A junction is represented by a node; physically a node can be a busbar, a joint, or other similar component. This requires no extra modeling, since the resistance and other electric effects are negligible [11].

Line

A line is modelled as a branch with a specified impedance and shunt resistance. Lines connect two nodes that are rated for the same voltage. The lines are modelled with a π model as well [11].

$$Z_{\text{series}} = r + jx$$
$$Y_{\text{shunt}} = \frac{2\pi f c}{\tan\delta + j}$$

The total resistance can be calculated using $r_1 = \rho * \frac{L}{A}$, where ρ is the resistivity of copper, A is the area of the cable and L the length.

Source

The source is representing the external network with a Thévenin's equivalence; it reduces the complex linear circuits back to a simple circuit with an impedance, and a voltage source. It has an infinite voltage source with an internal impedance, which simulates the behaviour of the higher voltage network responsible for the power delivery of the transformer [11].

Load

The load represents the power consumption of a customer, connected to the transformer. Customers will have a active and reactive power consumption. Both active power P and reactive power Q are constant during each timestep [11].

3-3-3 Stage 3: Power Flow Calculations

With the network topology established (Stage 1) and the mathematical models for all components assembled into the Y_{bus} matrix (Stage 2), the final stage is to execute the simulation and validate its output against real-world measurements. The `power-grid-model` library is used to solve the system of non-linear power flow equations corresponding to the intervals of the available power quality meter data.

Power Flow Calculation

The core of the simulation is solving the set of non-linear power flow equations (from Eq. 3-3), which can be represented abstractly as:

$$f(x) = y \tag{3-6}$$

where x is the vector of unknown voltage angles and magnitudes, $x = [\delta, |V|]^T$, and y is the vector of specified active and reactive power injections, $y = [P, Q]^T$. The `power-grid-model` library employs the robust Newton-Raphson method to solve this system iteratively. Starting from an initial guess $x(i)$ (a "flat start" where all $|V| = 1.0$ p.u. and $\delta = 0$), the algorithm performs the following steps until convergence:

Algorithm 2 Newton-Raphson Power Flow Solution.

Input: System Admittance Matrix Y_{bus}; Specified power vector $y_{\text{specified}} = [P, Q]^T$; Convergence tolerance ϵ.
Output: Solved voltage vector $x = [\delta, |V|]^T$; Slack bus power S_{slack}.
▷ ***Part 1: Initialization*** ◁
Initialize voltage vector $x(0)$ (e.g., "flat start": $|V_i| \leftarrow 1.0, \delta_i \leftarrow 0$ for all i).
Set iteration counter $k \leftarrow 0$.
▷ ***Part 2: Iterative Solution*** ◁
repeat
▷ *Step 1: Calculate Mismatch* ◁
$y(k) \leftarrow f(x(k))$ ▷ *Calculate power from current voltage $x(k)$ using Eq. 3-3*
$\Delta y(k) \leftarrow y_{\text{specified}} - y(k)$
▷ *Step 2: Formulate Jacobian* ◁
$J(k) \leftarrow \left[\frac{\partial f}{\partial x}\right]_{x=x(k)}$ ▷ *Build Jacobian matrix at current state $x(k)$*
▷ *Step 3: Solve Linear System* ◁
Solve $J(k)\Delta x(k) = \Delta y(k)$ for $\Delta x(k)$ ▷ *Typically via LU decomposition*
▷ *Step 4: Update State* ◁
$x(k+1) \leftarrow x(k) + \Delta x(k)$
$k \leftarrow k + 1$
until $\max(|\Delta y(k-1)|) < \epsilon$ ▷ *Check if mismatch is within tolerance*
▷ ***Part 3: Final Calculation*** ◁
$x_{\text{solved}} \leftarrow x(k)$
Calculate $S_{\text{slack}} = P_{\text{slack}} + jQ_{\text{slack}}$ using x_{solved} and (3-3) at the slack bus.
return x_{solved}, S_{slack}.

This iterative process is repeated until the magnitude of the mismatch vector $\Delta y(i)$ falls below a predefined tolerance, indicating that a solution has been found. The resulting aggregate power calculated at the slack bus ($P_{\text{sim}}, Q_{\text{sim}}$) provides the simulated data stream for validation against the measured transformer data, as presented in Section 3.5.

3-4 Case Study: Secondary Substation D114, Bonaire

This modeling pipeline was put to the test in Kralendijk, Bonaire.

3-4-1 Available Data

To build and validate the model, several data sources from the utility (WEB Bonaire) were utilized. In August 2024, a power quality meter was installed at the target secondary substation (D114), measuring active, reactive, and apparent power over all three phases at a 10-minute interval. This measurement data provides the ground truth for validating the aggregate load of the final model. For this case study, a one-week subset of this data is used.

A GIS is also available, registering most of the distribution assets. However, this dataset is incomplete and not always spatially accurate; links between customer connections and distribution lines are often missing or imprecise. To remedy this, a 2024 study was conducted by the distribution team in collaboration with NTCS Energy to manually notate the connections between *house connections* and their corresponding *transformers or distribution boxes*. This meter-transformer relationship data is a critical input for the topology reconstruction.

3-4-2 Topology Reconstruction

The foundational step of the modeling pipeline is the reconstruction of the grid's topology. This is a significant challenge as the available GIS data is incomplete, and the explicit connections between customers and distribution lines are often missing. To solve this, this study implements the optimization algorithm from Vassallo et al. [10], which is designed to identify customer topological paths using only static GIS data and known customer-to-transformer assignments.

Hypothetical Path Generation (A*) Given the large number of elements in the envelope, generating every possible hypothetical path ($\mathcal{H}$) is computationally impractical [10]. Therefore, the A* pathfinding algorithm was used to efficiently generate a viable subset of hypothetical paths ($\mathcal{H}^{I'}$) for each customer.

For each customer c known to be connected to D114, the A* algorithm searched for the most plausible paths from the customer's location to the D114 terminal junction. This search was constrained to use only the existing elements in the GIS database and to respect certain paramteres, such as a maximum distance for a connection.

Path Selection (ILP Optimization) The set of plausible paths from the A* algorithm was then fed into the Integer Linear Programming (ILP) optimizer. The optimizer's goal is to

(a) Original GIS data from WEB Bonaire **(b)** Estimated network topology

Figure 3-2: The estimated topology from the original data. Secondary substation D114 is colored Green

select one "estimated path" ($\hat{\mathcal{P}}$) for each customer from their list of hypothetical paths that best approximates the unknown "real path" ($\mathcal{P}$) [10].

The ILP problem, defined by the objective function in Equation (3-2), works to:

- **Maximize** the number of customers successfully connected to their correct terminal (D114) [10].
- **Enforce** constraints, such as ensuring that each customer is assigned only one path and that each network element is used by only one terminal, thereby guaranteeing a radial network structure [10].

Results The output of the ILP algorithm is the estimated topology for the D114 subnetwork, as visualized in Figure 3-2 in green. This reconstructed network serves as the foundational structure for the power flow model.

In total, 60 of the 64 customers known to be connected to D114 were successfully mapped by the algorithm. The four remaining connections could not be mapped. A diagnostic check revealed that these customers were located too far from the nearest "line" or "junction" element in the GIS database, making it impossible for the A* algorithm to find any hypothetical path that satisfied the distance constraints [10].

3-4-3 Model Parameterization

The generic grid component models described in the previous section were parameterized using the available data for substation D114.

Transformer

The main transformer parameters were derived from its specification plate, as listed in Table 3-2.

Table 3-2: Parameters for Transformer D114

Parameter	Value
u_1 (Primary Voltage)	12120 V
u_2 (Secondary Voltage)	400 V
s_n (Rated Power)	160 kVA
u_k (Short-circuit Voltage)	4.07 %
p_k (Copper Losses)	-
i_0 (No-load Current)	0.60 %
p_0 (Iron Losses)	- W
$winding_{from}$	2 (Delta)
$winding_{to}$	1 (Wye)
$clock$	11
tap_{side}	0 (HV)
tap_{min}	-2
tap_{max}	2
tap_{size}	303 V
tap_{pos}	0

A primary challenge was the absence of datasheet values for the transformer's loss parameters: p_k (short-circuit copper losses) and p_0 (no-load iron losses). An initial estimate for i_0 (no-load current) of 0.60% was used. To address this uncertainty, a sensitivity analysis was conducted to test the model's robustness and find the optimal parameterization.

First, a physically plausible range for the no-load parameters was established based on transformer literature [76]. This literature suggests i_0 is typically 1-2% of the rating, and p_0 is less than 1% of the total VA rating (i.e., < 1600W for this transformer). The model was found to be robust, with the final RMSE remaining stable across this entire physical range.

Lines

Line parameters were estimated based on standard conductor properties, as specific cable types were not available in the GIS data. In collaboration with the Engineers of WEB Bonaire, the reactance was determined using the physical parameters of the line setup geometry.

The positive-sequence reactance x_1 (in Ω/m) is given by Equation 3-7:

$$x_1 = \omega \left(\frac{\mu_0}{2\pi} \cdot \log \frac{GMD}{GMR} \right) \tag{3-7}$$

where ω is the angular frequency (rad/s), μ_0 is the magnetic constant, $GMR \approx e^{-0.25} r$ is the Geometric Mean Radius (with r being the conductor radius), and $GMD \approx (d_{l_1-l_2} \cdot d_{l_2-l_3} \cdot d_{l_1-l_3})^{1/3}$ is the Geometric Mean Distance between phases [77].

The positive-sequence capacitance c_1 (in F/m) is obtained using Equation 3-8:

$$c_1 = \frac{2\pi\epsilon_0}{\log(GMD/r)} \tag{3-8}$$

where ϵ_0 is the permittivity of free space. The resistance r_1 (in Ω/m) is obtained using $r_1 = \rho_{cu}/A$, where ρ_{cu} is the resistivity of copper and A is the cross-sectional area [78].

Loads

As customer-level smart meter data was unavailable, synthetic load profiles were generated. A normalized load profile was first derived from the original aggregate power quality meter data. This profile was then scaled by the average power consumption for each of the 64 customers.

To introduce realistic variation, each data point in the scaled profiles was used as the mean μ for a normal distribution. A new value was then drawn from this distribution using a standard deviation (σ) equal to 25% of the mean load at that timestep. This method creates unique, stochastic profiles for each customer while preserving the "slow" characteristics (i.e., the daily curve) of the original aggregate load.

External Grid (Source)

The external grid is represented as a source component. For this study, the default values are used: $u_{\text{ref-angle}} = 0$, $s_k = 1 \times 10^{10}$ VA, $r/x = 0.1$, and $z_0/z_1 = 1$. Since the transformer is stepping down the voltage, $u_{\text{ref}} = 12$ kV.

3-5 Powerflow Calculation Results and analysis

Following the parameterization of all grid assets, the complete power flow simulation was executed. The model's aggregated active and reactive power output at the transformer was validated against the ground truth measurement data from the power quality meter.

A one-week period from September, a time characterized by high demand, was used for the validation. The simulation, compensating for the 60 of 64 connected loads, produced a strong fit with the measured data, as shown in Figure 3-3. The final error estimates are as follows:

- Active power RMSE: 6.08%
- Reactive power RMSE: 5.74%

When performing a crosscorrelation analysis, it can be concluded that the error is significantly zero mean white noise, which is introduced in the synthetic load profile generation. This analysis provides crucial insights into both overall transformer loading and the calculated loading of individual lines. However, a limitation of this study is the inability to validate the specific line loading results, as no granular measurements were available for these segments. The overall power flow analysis is accurate, but its precision could be enhanced in the future by incorporating real customer-level load profiles and more precise electrical parameters for both the lines and the transformer.

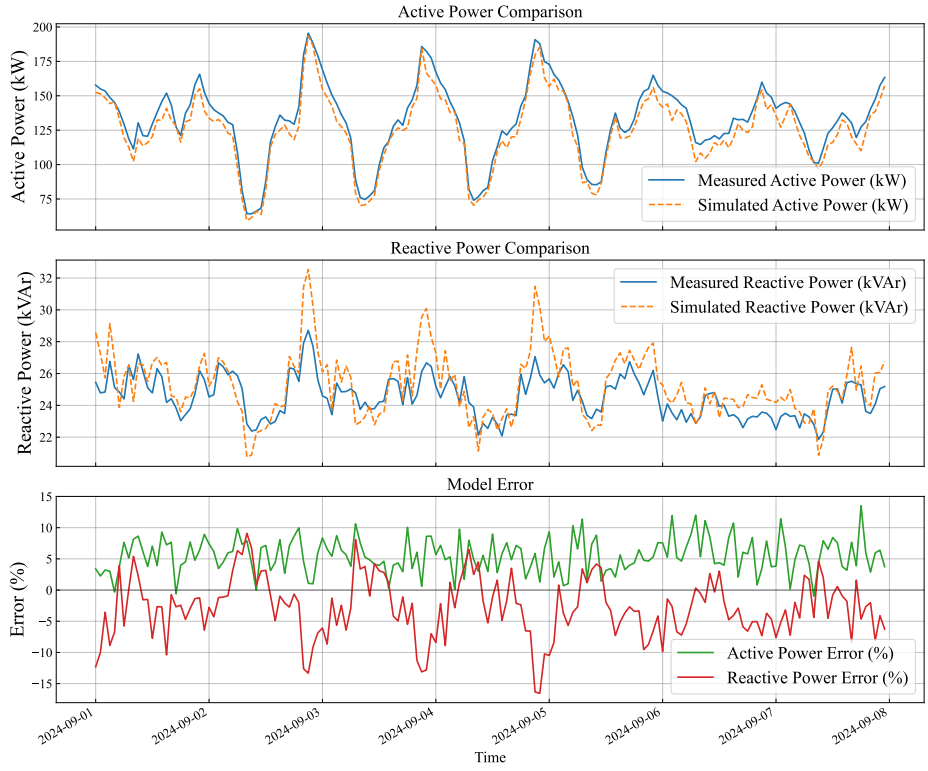

Figure 3-3: Comparison of Simulated and Measured Active and Reactive Power for one week in September.

3-5-1 Discussion

The power flow analysis, achieving an RMSE of $\approx 6\%$, validates the pipeline as a viable method for modeling a limited-data grid. The primary insights, however, come from analyzing the model's limitations.

While most parameters in this study are based on the actual physical parameters of the assets, there are still some estimations to be made. This includes power line quality, transformer efficiency and losses, or load profile estimations. While the topology estimation algorithm performs well, there are still 4 missing loads. For larger areas, this can become a accumulating error factor. Verifying the load profile dynamics with actual smart meter data will significantly improve the accuracy of the prediction.

3-5-2 Conclusion and future work

This paper presented a complete, data-driven pipeline for modeling low-voltage distribution grids where traditional data sources like smart meters and accurate GIS data are limited. By combining an ILP-based topology reconstruction method with a standard power flow model, a viable representation of a 160kVA grid in Bonaire was created.

The model was validated against real-world power quality meter data, achieving a low aggregate error (RMSE of $\approx 6\%$). The contribution of this work is to show that a combination of existing techniques can provide high quality modeling solution which can provide insight for control and investment decisions, based on the available data. The first and most critical step is to replace the synthetic load profiles with real smart meter data, if it becomes available. This would isolate the error contributions from the topology and component models.

Second, the robustness of the ILP-derived topology should be tested. As suggested in the project notes, conducting cross-validation by applying this entire pipeline to other transformers in the distribution network would be a powerful test of the method's generalizability.

Finally, while the current model provides a strong physical representation, machine learning techniques, such as neural networks, could be explored to refine parameter estimation or capture complex load behaviors not represented by the current statistical load model.

Chapter 4

Conclusions and Recommendations

4-1 Conclusion

This thesis set out to address two primary modeling challenges for integrating Hybrid Energy Storage (HES) systems into modern power grids. The research successfully delivered on both objectives, providing a validated toolset for analyzing the role of AEM electrolyzers in providing distribution grid support.

The main conclusions are twofold. First, a high-fidelity, control-oriented hybrid model for an AEM electrolyzer was successfully developed and validated. As detailed in Chapter 2, a conventional physics-based Equivalent Circuit Model (ECM) was insufficient (26.00% fit). By combining the interpretable RC model with a lightweight neural network trained to correct for thermal dynamics, the resulting hybrid model achieved a 92.03% fit. This model balances physical interpretability, computational efficiency, and high accuracy, making it ideal for advanced, model-based control applications.

Second, a complete and viable data-driven pipeline for modeling limited-data LV distribution grids was demonstrated. As presented in Chapter 3, a reliable "digital twin" of a 160kVA substation in Bonaire was constructed despite incomplete GIS data. By combining an ILP algorithm for topology reconstruction with a standard power flow model, the simulation was validated against aggregate power quality meter data, achieving a low RMSE of approximately 6% for both active and reactive power. This contribution validates that a combination of existing techniques can produce a reliable grid model for DSOs, even in data-scarce environments.

In synthesis, this thesis has successfully developed the two core components required to study the interaction between emerging HES assets and LV grids: a robust model for the AEM electrolyzer asset and a reliable model for the distribution grid it connects to.

4-2 Recommendations

Based on the findings and models developed in this thesis, several avenues for future research are recommended.

Integrated system co-simulation: the most exciting and logical next step is to perform an integrated system co-simulation by integrating the AEM electrolyzer model (Chapter 2) into the LV grid model (Chapter 3). This would enable a detailed analysis of the electrolyzer's impact on the grid's voltage profile, phase balance, and line congestion in its intended grid-support role.

Advancements in electrolyzer modeling and control: For future research, the computationally efficient hybrid model should be utilized within a Model Predictive Control (MPC) framework for real-time optimization. Furthermore, incorporating online parameter estimation (e.g., Kalman filters) would allow the model to adapt to long-term stack degradation, while re-validating the model with high-frequency data would allow for grid stability studies.

To enhance the distribution grid model, the largest source of uncertainty, the synthetic load profiles, should be replaced with real smart meter data if it becomes available. In its absence, more advanced machine learning techniques (e.g., GANs) could create more realistic profiles. Using the existing pipeline, a cross validation against other measured transformers or substation can be performed. This was not possible at the time of research due to lack of measurements.

Finally, a holistic HES and techno-economic analysis should be performed. This involves expanding the HES model to include the full Power-to-Hydrogen-to-Power (P2H2P) cycle (i.e., hydrogen storage and a fuel cell) and conducting a comprehensive techno-economic study to evaluate the business case for using AEM electrolyzers for grid support in locations like Bonaire.

Bibliography

[1] PARIS AGREEMENT. Technical report.

[2] N. Phuangpornpitak and S. Tia. Opportunities and challenges of integrating renewable energy in smart grid system. *Energy Procedia*, 34:282–290, 2013.

[3] J. Mitali, S. Dhinakaran, and A. A. Mohamad. Energy storage systems: a review, 9 2022.

[4] Salman Hajiaghasi, Ahmad Salemnia, and Mohsen Hamzeh. Hybrid energy storage system for microgrids applications: A review, 2 2019.

[5] John Newman and Nitash P Balsara. *Electrochemical Systems*. John Wiley & Sons, 2021.

[6] S. Shiva Kumar and Hankwon Lim. An overview of water electrolysis technologies for green hydrogen production, 11 2022.

[7] Nikolaos Damianakis, Gautham Ram Chandra Mouli, Pavol Bauer, and Yunhe Yu. Assessing the grid impact of Electric Vehicles, Heat Pumps & PV generation in Dutch LV distribution grids. *Applied Energy*, 352, 12 2023.

[8] Yanhui Xu and Zilin Deng. Bi-level planning of microgrid considering seasonal hydrogen storage and efficiency degradation of electrolyzer. *IEEE Transactions on Industry Applications*, 2024.

[9] Xin Lin and Ramon Zamora. Controls of hybrid energy storage systems in microgrids: Critical review, case study and future trends, 3 2022.

[10] Vassallo Maurizio, Leerschool Adrien, Bahmanyar Alireza, Laurine Duchesne, Gerard Simon, Wehenkel Thomas, and Ernst Damien. An Optimization Algorithm for Customer Topological Paths Identification in Electrical Distribution Networks. Technical report.

[11] Yu Xiang, Peter Salemink, Werner van Westering, Nitish Bharambe, Martinus G.H. Govers, Jonas van den Bogaard, Bram Stoeller, Zhen Wang, Jerry Jinfeng Guo, Santiago Figueroa Manrique, Laurynas Jagutis, Chenguang Wang, Marc van Raalte, and Contributors to the LF Energy project Power Grid Model. PowerGridModel/power-grid-model.

[12] Teresa Mendiara, Francisco García-Labiano, Alberto Abad, Pilar Gayán, Luis F de Diego, Maria Teresa Izquierdo, and Juan Adánez. Negative CO_2 emissions through the use of biofuels in chemical looping technology: A review. *Applied energy*, 232:657–684, 2018.

[13] United Nations. The Paris Agreement, 2015.

[14] Emmanuel Ejuh Che, Kang Roland Abeng, Chu Donatus Iweh, George J Tsekouras, and Armand Fopah-Lele. The Impact of Integrating Variable Renewable Energy Sources into Grid-Connected Power Systems: Challenges, Mitigation Strategies, and Prospects. *Energies*, 18(3):689, 2025.

[15] Massiagbe Fatoumata Diabate, Harish Sarma Krishnamoorthy, and Jian Shi. Optimal design and modeling of a hybrid energy storage system including battery and hydrogen in dc microgrids. *IEEE Transactions on Industry Applications*, 61(5):7571–7583, 2025.

[16] Mohamed R. Elkadeem, Kotb M. Kotb, Atif S. Alzahrani, and Mohammad A. Abido. Design and global sensitivity analysis of a power-to-hydrogen-to-power-based multi-energy microgrid under uncertainty. *IEEE Transactions on Industry Applications*, 61(2):1811–1827, 2025.

[17] Mostafa El-Shafie. Hydrogen production by water electrolysis technologies: A review. *Results in Engineering*, 20, 12 2023.

[18] Aaquib Firdous, Chandra Prakash Barala, Parul Mathuria, Rohit Bhakar, and Mohammad Shahidehpour. Utility-scale green hydrogen system's operational flexibility under temperature dynamics. *IEEE Transactions on Industry Applications*, pages 1–11, 2025.

[19] Mohit Kumar, Nandhakumar Vijayakumar, Kevin Pretorius, and Terry Tadlock. Reliable and efficient high power DC distribution for hydrogen electrolysis. *IEEE Transactions on Industry Applications*, 2025.

[20] Sooraj Suresh Kumar, Nirmal Mukundan C. M., and Jayaprakash P. Modified lms control for a grid interactive pv–fuel cell–electrolyzer hybrid system with power dispatch to the grid. *IEEE Transactions on Industry Applications*, 58(6):7907–7918, 2022.

[21] Rammohan Rao Makineni, Ashish P. Agalgaonkar, Kashem M. Muttaqi, Md. Rabiul Islam, and Danny Sutanto. Integral sliding mode control of a stacked interleaved buck converter for electrolyzers supplied with renewable energy sources. *IEEE Transactions on Industry Applications*, 61(1):450–462, 2025.

[22] Arta Mohammad-Alikhani, Amin Mahmoudi, Rahmat Khezri, and Solmaz Kahourzade. Multiobjective optimization of system configuration and component capacity in an ac minigrid hybrid power system. *IEEE Transactions on Industry Applications*, 58(3):4158–4170, 2022.

[23] Mohamed Khalid Ratib, Kashem M Muttaqi, Md Rabiul Islam, Danny Sutanto, and Ashish P Agalgaonkar. Electrical circuit modeling of proton exchange membrane electrolyzer: The state-of-the-art, current challenges, and recommendations. *International journal of hydrogen energy*, 49:625–645, 2024.

[24] Rafal Bernat, Jaroslaw Milewski, Olaf Dybinski, Aliaksandr Martsinchyk, and Pavel Shuhayeu. Review of AEM electrolysis research from the perspective of developing a reliable model. *Energies*, 17(20):5030, 2024.

[25] Mingzhi He, Gongzhe Nie, Haoran Yang, Binghui Li, Shuhan Zhou, Xiongzheng Wang, and Xin Meng. A generic equivalent circuit model for PEM electrolyzer with multi-timescale and stages under multi-mode control. *Applied energy*, 359:122728, 2024.

[26] Álvaro Iribarren, David Elizondo, Ernesto L Barrios, Harkaitz Ibaiondo, Alain Sanchez-Ruiz, Joseba Arza, Pablo Sanchis, and Alfredo Ursúa. Dynamic modeling of a pressurized alkaline water electrolyzer: A multiphysics approach. *IEEE Transactions on Industry Applications*, 59(3):3741–3753, 2023.

[27] M. Ihsan Karamangil Merve Tekin. Development of dual polarization battery model with high accuracy for a lithium-ion battery cell under dynamic driving cycle conditions. *Heliyon*, 10(7):e28454, March 2024. Describes Rint model as simplest ECM with only ohmic resistance, unable to capture polarization.

[28] Development of a fast running equivalent circuit model with thermal predictions for battery management applications. *Batteries*, 10(6):215, June 2024. First-order RC model for fast dynamics, balancing accuracy and computational efficiency.

[29] Md Biplob Hossain, Md Rabiul Islam, Kashem M Muttaqi, Danny Sutanto, and Ashish P Agalgaonkar. Dynamic electrical circuit modeling of a proton exchange membrane electrolyzer for frequency stability, resiliency, and sensitivity analysis in a power grid. *IEEE Transactions on Industry Applications*, 59(6):7271–7281, 2023.

[30] Minho Kim et al. Online identification of lithium-ion battery model parameters with initial value uncertainty and measurement noise. *Chinese Journal of Mechanical Engineering*, January 2023. Thevenin model captures primary dynamics without heavy computational cost.

[31] Diffusion-equation-based electrical modeling for high-power lithium titanium oxide batteries. *Batteries*, 10(7):238, July 2024. Second-order RC with separate branches for electrochemical and concentration polarization.

[32] Xiuliang Zhao, Yanlong Liu, Zhengyu Yang, Ruochen Wang, Liang Liu, Limei Wang, and Yun Wang. A modified high c-rate battery equivalent circuit model based on current dependence and concentration modification. *Electrochimica Acta*, 478:143833, 2024.

[33] A. Alshawabkeh, M. Matar, and F. Almutairy. Parameters identification for lithium-ion battery models using the levenberg–marquardt algorithm. *World Electric Vehicle Journal*, 15(9):406, September 2024. N-order Thevenin model, discusses computational efficiency vs. accuracy trade-offs.

[34] Online reduced complexity parameter estimation technique for equivalent circuit model of lithium-ion battery. *Electric Power Systems Research*, May 2020. Multi-RC models have high computational cost for real-time applications, discusses identifiability challenges.

[35] Su et al. State-of-health estimation of lithium-ion batteries: A comprehensive literature review from cell to pack levels. *Energy Conversion and Economics*, August 2024. Discusses ECM complexity, first- vs second-order RC trade-offs.

[36] Sunil Subedi, Yonghao Gui, and Yaosuo Xue. Applications of data-driven dynamic modeling of power converters in power systems: An overview. *IEEE Transactions on Industry Applications*, 61(2):2434–2456, March 2025. Special Issue: Convergence of Data-driven and Physics-based Approaches in Power System Analysis, Optimization, and Control.

[37] Tao Wang, Jinyi Wang, Chang Zhang, Pengjie Wang, Zhibo Ren, Haijiao Guo, Zhan Wu, and Fan Wang. Direct operational data-driven workflow for dynamic voltage prediction of commercial alkaline water electrolyzers based on artificial neural network (ANN). *Fuel*, 376:132624, 2024.

[38] Islam Zerrougui, Zhongliang Li, and Daniel Hissel. Physics-informed neural network for modeling and predicting temperature fluctuations in proton exchange membrane electrolysis. *Energy and AI*, 20:100474, 2025.

[39] Jodel Cornelio, Syamil Mohd Razak, Young Cho, Hui-Hai Liu, Ravimadhav Vaidya, and Behnam Jafarpour. Residual learning to integrate neural network and physics-based models for improved production prediction in unconventional reservoirs. *SPE Journal*, 27(06):3328–3350, 2022.

[40] Zhen Chen and Dongbin Xiu. On generalized residual network for deep learning of unknown dynamical systems. *Journal of Computational Physics*, 438:110362, 2021.

[41] Syamil Mohd Razak, Jodel Cornelio, Young Cho, Hui-Hai Liu, Ravimadhav Vaidya, and Behnam Jafarpour. A dynamic residual learning approach to improve physics-constrained neural network predictions in unconventional reservoirs. In *SPE Middle East Oil and Gas Show and Conference*, page D021S084R005. SPE, 2023.

[42] Bo Han, Stuart M Steen III, Jingke Mo, and Feng-Yuan Zhang. Electrochemical performance modeling of a proton exchange membrane electrolyzer cell for hydrogen energy. *International Journal of Hydrogen Energy*, 40(22):7006–7016, 2015.

[43] Khaled Lawand, Suhas Nuggehalli Sampathkumar, Zoé Mury, and Jan Van Herle. Membrane electrode assembly simulation of anion exchange membrane water electrolysis. *Journal of Power Sources*, 595:234047, 2024.

[44] Damien Guilbert and Gianpaolo Vitale. Dynamic emulation of a PEM electrolyzer by time constant based exponential model. *Energies*, 12(4), 2 2019.

[45] Md Biplob Hossain, Md Rabiul Islam, Kashem M. Muttaqi, Danny Sutanto, and Ashish P. Agalgaonkar. Dynamic Electrical Equivalent Circuit Modeling of the Grid-Scale Proton Exchange Membrane Electrolyzer for Ancillary Services. In *Conference Record - IAS Annual Meeting (IEEE Industry Applications Society)*, volume 2022-October. Institute of Electrical and Electronics Engineers Inc., 2022.

[46] Mustafa Ergin Şahin. An overview of different water electrolyzer types for hydrogen production. *Energies*, 17(19):4944, 2024.

[47] Lennart Ljung. *System identification: Theory for the user*. Prentice Hall, 1999.

[48] Mohammad Khosravi and Roy S Smith. Nonlinear system identification with prior knowledge on the region of attraction. *IEEE Control Systems Letters*, 5(3):1091–1096, 2021.

[49] Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.

[50] Lennart Ljung, Carl Andersson, Koen Tiels, and Thomas B Schön. Deep learning and system identification. *IFAC-PapersOnLine*, 53(2):1175–1181, 2020.

[51] Mohammad Khosravi. Representer theorem for learning Koopman operators. *IEEE Transactions on Automatic Control*, 2023.

[52] Mohammad Khosravi and Roy S Smith. Convex nonparametric formulation for identification of gradient flows. *IEEE Control Systems Letters*, 5(3):1097–1102, 2021.

[53] L. An, T. S. Zhao, Z. H. Chai, P. Tan, and L. Zeng. Mathematical modeling of an anion-exchange membrane water electrolyzer for hydrogen production. *International Journal of Hydrogen Energy*, 39(35):19869–19876, 12 2014.

[54] Enapter. Electrolyser EL 2.1. Technical report.

[55] Erwan Tardy, Yann Bultel, Florence Druart, Antoine Bonnefont, Melaine Guillou, and Benoit Latour. Three-Dimensional Modeling of Anion Exchange Membrane Electrolysis: A Two-Phase Flow Approach. *Energies*, 17(13), 7 2024.

[56] Jay T. Pukrushpan, Huei Peng, and Anna G. Stefanopoulou. Control-oriented modeling and analysis for automotive fuel cell systems. *Journal of Dynamic Systems, Measurement and Control, Transactions of the ASME*, 126(1):14–25, 2004.

[57] Fatemeh Homayouni, Ramin Roshandel, and Ali Asghar Hamidi. Sizing and performance analysis of standalone hybrid photovoltaic/battery/hydrogen storage technology power generation systems based on the energy hub concept. *International Journal of Green Energy*, 14(2):121–134, 1 2017.

[58] C. Rozain and P. Millet. Electrochemical characterization of Polymer Electrolyte Membrane Water Electrolysis Cells. *Electrochimica Acta*, 131:160–167, 6 2014.

[59] Hamish Andrew Miller, Karel Bouzek, Jaromir Hnat, Stefan Loos, Christian Immanuel Bernäcker, Thomas Weißgärber, Lars Röntzsch, and Jochen Meier-Haack. Green hydrogen from anion exchange membrane water electrolysis: A review of recent developments in critical materials and operating conditions, 5 2020.

[60] Benjamin Flamm, Christian Peter, Felix N. Büchi, and John Lygeros. Electrolyzer modeling and real-time control for optimized production of hydrogen gas. *Applied Energy*, 281, 1 2021.

[61] Shing Cheng Chang, Ru En Gu, and Yen Hsin Chan. Parameter Analysis of Anion Exchange Membrane Water Electrolysis System by Numerical Simulation. *Energies*, 17(22), 11 2024.

[62] Tohid Adibi, Atta Sojoudi, and Suvash C. Saha. Modeling of thermal performance of a commercial alkaline electrolyzer supplied with various electrical currents. *International Journal of Thermofluids*, 13, 2 2022.

[63] Abu Bakr Pengwah, Luxin Fang, Reza Razzaghi, and Lachlan L.H. Andrew. Topology Identification of Radial Distribution Networks Using Smart Meter Data. *IEEE Systems Journal*, 16(4):5708–5719, 12 2022.

[64] Bartłomiej Mroczek and Paweł Pijarski. Dso strategies proposal for the lv grid of the future. *Energies*, 14(19), 10 2021.

[65] Caio O. Pereira, Vinícius C. Cunha, Tiago R. Ricciardi, Victor B. Riboldi, and Ji Tuo. *Pre-Installation Studies of a BESS in a Real LV Network with High PV Penetration.* IEEE, 2019.

[66] Guido Cavraro and Vassilis Kekatos. Graph algorithms for topology identification using power grid probing. *IEEE Control Systems Letters*, 2(4):689–694, 10 2018.

[67] Wided Medjroubi, Ulf Philipp Müller, Malte Scharf, Carsten Matke, and David Kleinhans. Open Data in Power Grid Modelling: New Approaches Towards Transparent Grid Models. *Energy Reports*, 3:14–21, 11 2017.

[68] Mohammad Farajollahi, Alireza Shahsavari, and Hamed Mohsenian-Rad. Topology Identification in Distribution Systems Using Line Current Sensors: An MILP Approach. *IEEE Transactions on Smart Grid*, 11(2):1159–1170, 3 2020.

[69] Xing He, Robert C. Qiu, Qian Ai, and Tianyi Zhu. A Hybrid Framework for Topology Identification of Distribution Grid with Renewables Integration. *IEEE Transactions on Power Systems*, 36(2):1493–1503, 3 2021.

[70] Yang Weng, Yizheng Liao, and Ram Rajagopal. Distributed Energy Resources Topology Identification via Graphical Modeling. *IEEE Transactions on Power Systems*, 32(4):2682–2694, 7 2017.

[71] Haipeng Zhang, Jian Zhao, Xiaoyu Wang, and Yi Xuan. Low-Voltage Distribution Grid Topology Identification With Latent Tree Model. *IEEE Transactions on Smart Grid*, 13(3):2158–2169, 5 2022.

[72] Giuliano Andrea Pagani and Marco Aiello. The Power Grid as a complex network: A survey. *Physica A: Statistical Mechanics and its Applications*, 392(11):2688–2700, 6 2013.

[73] Deepjyoti Deka, Vassilis Kekatos, and Guido Cavraro. Learning Distribution Grid Topologies: A Tutorial. *IEEE Transactions on Smart Grid*, 15(1):999–1013, 1 2024.

[74] Changbin Hu, Yang Wang, Shanna Luo, and Fudong Zhang. State-space model of an inverter-based micro-grid. In *IGBSG 2018 - 2018 International Conference on Intelligent Green Building and Smart Grid*, pages 1–7. Institute of Electrical and Electronics Engineers Inc., 6 2018.

[75] LF Energy. Power Grid Model. https://lfenergy.org/projects/power-grid-model/, oct 2025. Accessed: November 10, 2025.

[76] Rejean Arseneau, Eddy So, and Ernst Hanique. Measurements and correction of no-load losses of power transformers. *IEEE Transactions on Instrumentation and Measurement*, 54(2):503–506, 4 2005.

[77] John Grainger and William Stevenson. *Power Systems Analysis*. Power & Energy. McGraw-Hill Publishing, London, England, jun 1994.

[78] Dr. Mithun Mondal. Capacitance Calculation Mastery: Transmission Line Insights — engineeringdevotion.com. https://www.engineeringdevotion.com/electric-power-systems/solved-problems/transmission-line-capacitance.html. [Accessed 10-11-2025].

Nomenclature

List of Acronyms

AEM	Anion Exchange Membrane
DER	Distributed Energy Resources
DNN	Deep Neural Network
DSO	Distribution Service Operators
ECM	Equivalent Circuit Models
ESS	Energy Storage Systems
HES	Hybrid Energy Storage
ILP	Integer Linear Programming
LV	Low Voltage
NN	Neural Networks
P2H2P	Power-to-Hydrogen-to-Power
PEM	Proton Exchange Membrane
PINN	Physics-Informed Neural Networks
RC	Resistance-Capacitance
RES	Renewable Energy Sources
TPI	Topological Path Identification

Acknowledgements

First and foremost, I want to thank Dr. ir. Mohammad Khosravi. This project would not have been possible without you, and the opportunity to work with such a dedicated supervisor has been a blast. Thank you for your insights and our long discussions on modeling, entrepreneurship, and politics. You motivated me to aim for a first publication; we will see the results of that later this year. My thanks also go to Dr. Ir. Amirreza Silani for your help with the electrolyzer modeling, your sharp contributions to our paper, and our general discussions on modeling and power markets.

Thank you to the team at The Green Village for welcoming me and helping me with all the necessary data and information. I'd like to highlight Joep, who helped me gain a thorough understanding of the Energy Hub and was a great sounding board for new ideas. Also, Daan and Emma, thank you for the fun times while writing our theses together.

I was also fortunate to work with the enthusiastic and dedicated people at NTCS Energy. Jan, thank you for all your support over the years. I couldn't imagine a better place than Bonaire to work side-by-side with you on the grid modeling studies. My thanks also go to Jelmer and Mark, who are crazy enough to help make projects like these possible. To all the people from WEB Bonaire, especially Gerrit: thank you for inviting and hosting me on your beautiful island.

And then, my friends and family. Pap, mam, thank you for always supporting me during my studies. You're the best parents I could hope for! To my siblings and their other halves, Sanne en Lindert, Joren en Babette, and Elja: you all set the bar way too high! Thank you, family Zeinstra, for always being so interested and supportive. Studying S&C with Tom and Joshua made it a lot more fun. Thank you for all the study sessions, discussions, and healthy competition; I'm looking forward to our next chapter! To my housemates, Pepe, Jasper, Luc, Hidde en Foef: thanks for making the house I came back to every day feel like home this past year. For all the needed distractions, I knew I could count on Jasper Smit, Jim, Sander, Thomas, Gerben, Vince, Coen, Sophie, Nadia, Pepijn, Stephanie and the many more I might be forgetting.

And to the person who supported me most all those years: Dear Kira, thank you for always being there for me. I'm so proud of what you delivered this year with your own graduation, all while giving me all the support I could wish for. Together we can build the world!

Delft, University of Technology — Ruben van der Horst
November 11, 2025