Department of Electrical, Electronic, and
Information Engineering
"Guglielmo Marconi" – DEI

Dipartimento di Ingegneria
dell'Informazione,
Elettrica e Matematica Applicata

ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA

CURRICULUM – SOLAR INTERMITTENCY AND STORAGE

DOCTORAL DISSERTATION

Forecasting and Optimisation of PV–Battery Hybrid Systems: Methods for Advanced Renewable Integration

Presented by:
Costanza Luppi

Supervisor:
Prof. Vincenzo Cirimele

Co-supervisors:
Prof. Mattia Ricco

Final Exam Year – 2025

Dottorato di Interesse Nazionale
"PHOTOVOLTAICS"

Curricula:

- Solar cells technologies and lifecycle
- Design and integration
- Monitoring and diagnosis
- Power electronics and control
- Solar intermittency and storage
- Distributed generation and grid connection

Acknowledgements

I would like to express my deepest gratitude to Professor Gabriele Grandi for welcoming me into his wonderful research group and for believing in my potential despite my different academic background.

My sincere thanks go to Professor Vincenzo Cirimele and Professor Mattia Ricco, my supervisor and co-supervisor, for their constant guidance, patience, and support, and for the constructive dialogue that has profoundly enriched both my personal and professional growth.

I am deeply grateful to Professor Catia Arbizzani for her invaluable support and inspiring mentorship during the final stages of my Ph.D. journey. Thanks to her, I had the privilege of spending a remarkable period at the Fraunhofer Research Center, where I met talented and welcoming people who greatly enriched my academic and personal experience.

Special thanks go to Professor Vincenzo Nesi, the first professor I met in my first year at university. Thank you for being a fundamental guide throughout my personal and professional journey until the very last day, for helping me find my path, and for supporting me with warmth and trust at every stage of this journey.

My heartfelt appreciation goes to Professor Riccardo Mandrioli for his unwavering moral and emotional support throughout these years. You taught me what it truly means to do research, how to craft meaningful presentations, and, above all, reminded me of the importance of humor and joy along the way.

A special thank you to Paolo Pilati, my partner in crime, with whom I shared laughter, challenges, and countless unforgettable moments that made this Ph.D. journey truly special.

I am also deeply thankful to Sara Baldisserri for sharing with me both the joys and frustrations of this adventure, and for all those conversations over cigarettes that turned into reflections on life. Your friendship is a gift I will always treasure.

My gratitude extends to Riccardo Barbone, Lohith Kumar, Jiayi Jeng, Riccardo Adinolfi Borea, Gianluca Gentile, and Francesco Maria Tiburtini for their companionship and for all the moments of lightheartedness that brightened this journey.

A heartfelt thank you goes to my family. To my mother, for her unconditional love, for being my anchor and confidant through every decision and doubt, none of this would have been possible without you. To my sister Ludovica (Ica), who has patiently endured twenty years of my presentations, drafts, and revisions — you are, and always will be, life's most precious gift to me. To my father, whose kindness and humor have always managed to make me smile even on the hardest days, and whose support I will always cherish.

Finally, my love and gratitude to Lucrezia, Margherita, Benedetta, Flaminia, and Davide, for your affection, friendship, and for being a constant source of joy in my life.

Abstract

The energy transition is profoundly reshaping the electricity sector, with renewables at the forefront. Photovoltaic (PV) power is a key driver of decarbonisation thanks to its modularity, scalability, and falling costs. However, the variability of solar power complicates the balance between supply and demand, increases the risk of curtailment, and necessitates the use of ancillary services to ensure grid stability. Accurate PV forecasting is therefore essential to anticipate fluctuations and enable more reliable system operation.

In parallel with the growing share of renewables, the need for energy storage systems (ESS) has increased. Storage is crucial to mitigate the variability of PV and wind generation, ensuring continuity of supply and grid stability. While lithium-ion batteries dominate the market for their efficiency and maturity, vanadium redox flow batteries (VRFB) are emerging as a promising alternative, particularly for long-duration applications where the decoupling of power and energy provides greater flexibility. Optimising the design and sizing of storage is therefore essential to maximise performance and ensure economic feasibility.

This thesis develops an integrated framework that combines advanced PV forecasting with the development of a model for the optimisation of battery energy storage systems, focusing on VRFB technology. The two main objectives are: (i) the design of forecasting models based on gradient boosting (GBM) and LightGBM, enhanced with hybrid feature engineering strategies; and (ii) the development of an optimisation model tailored to VRFB technology for techno-economic assessment in hybrid renewable systems.

On the forecasting side, the proposed approach leverages ensemble learning to capture nonlinear meteorological patterns and is validated with historical and forecast weather

data, emphasizing robustness to input uncertainty. On the storage side, the optimisation model incorporates key VRFB characteristics such as decoupled power and energy capacity. Application to a PV–wind case study enables optimal sizing and benchmarking against lithium-ion batteries.

Results show that accurate PV forecasting enables more efficient use of renewable energy, while VRFBs prove particularly competitive in long-duration applications compared to lithium-ion batteries.

Overall, the integration of forecasting and optimisation emerges as a key factor not only for improving the techno-economic performance of renewable systems, but also for supporting their wider, sustainable and reliable deployment in future low-carbon energy networks.

Contents

Glossary

ANNs	Artificial Neural Networks
API	Application Programming Interfaces
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
ARX	Autoregressive with Exogenous Variables
BESS	Battery Energy Storage System
CNN	Convolutional Neural Network
CRF	Capital Recovery Factor
DHI	Diffuse Horizontal Irradiance
DNI	Direct Normal Irradiance
EML	Ensemble Machine Learning
ESS	Energy Storage System
GA	Genetic Algorithms

GB	Gradient Boosting
GBM	Gradient Boosting Machines
GHI	Global Horizontal Irradiance
GRU	Gated Recurrent Unit
IEA	International Energy Agency
k-NN	k-Nearest Neighbors
LightGBM	Light Gradient Boosting Machine
Li-ion	Lithium-ion
LP	Linear Programming
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MI	Mutual Information
MILP	Mixed-Integer Linear Programming
MINLP	Mixed-Integer Nonlinear Programming
ML	Machine Learning
MLP	Multilayer Perceptrons
NLP	Nonlinear Programming
NWP	Numerical Weather Prediction
PCA	Principal Component Analysis
PSO	Particle Swarm Optimization
PV	Photovoltaics

RF	Random Forests
RFB	Redox Flow Batteries
RFE	Recursive Feature Elimination
RMSE	Root-mean-square error
RNN	Recurrent Neural Networks
R^2	Coefficient of determination
SARIMA	Seasonal Autoregressive Integrated Moving Average
SARIMAX	Seasonal ARIMA with Exogenous Variables
SOC	State of Charge
SVM	Support Vector Machines
VAR	Vector Autoregression
VARX	Vector Autoregression with Exogenous Variables
VRFB	Vanadium Redox Flow Batteries
XAI	Explainable Artificial Intelligence
XGBoost	eXtreme Gradient Boosting

List of Symbols

α	Empirical coefficient in feature χ_3
β_1	Empirical coefficient in feature χ_1
β_2	Empirical coefficient in feature χ_2
χ_1	Synthetic irradiance feature
χ_2	Synthetic irradiance feature
χ_3	Synthetic irradiance feature
χ_4	Synthetic irradiance feature
χ_5	Synthetic irradiance feature
η_{ch}	Charging efficiency [-]
η_{dis}	Discharging efficiency [-]
γ	Empirical coefficient in feature χ_4
C	Total system cost [€]
c	Electrolyte cost including tanks [€/kWh]
c	Total cost [€/kWh]

Cl	Cloud cover [%]
e	Stack cost per installed power [€/kW]
f_t	Electricity demand during interval t [kWh]
G	Global irradiance [W/m^2]
g_t	Renewable generation during interval t [kWh]
$G_{\max}$	Maximum observed irradiance [W/m^2]
k	Number of installed stacks
p_t^{buy}	Buying electricity price [€/kWh]
p_t^{sell}	Selling electricity price [€/kWh]
P^{stk}	Rated power of one stack [kW]
p_t	Net electricity balance at interval t [kWh]
q_t^{in}	Energy imported from the grid during interval t [kWh]
q_t^{out}	Energy exported to the grid during interval t [kWh]
s_0	Initial state of charge [kWh]
s_t	State of charge at the end of interval t [kWh]
w_1	Weighting factor in feature χ_5
x_t^{ch}	Charging energy during interval t [kWh]
$x_t^{\text{dis,raw}}$	Raw discharging energy (before efficiency) [kWh]
x_t^{dis}	Effective discharging energy [kWh]
y	Total installed energy capacity [kWh]

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Thesis motivation and Objectives

The global energy transition is driving an extraordinary transformation of the electricity sector, with renewable energy sources at the forefront of this process. According to the International Energy Agency, in 2023 renewables accounted for more than 80% of new power capacity additions, with solar photovoltaics (PV) representing the largest share of this growth [1, 2]. This unprecedented expansion is closely tied to the decarbonisation objectives established at both global and regional levels, such as the European Green Deal and the EU Climate Law, which commit the European Union to achieving climate neutrality by 2050 [3–5]. In this framework, PV energy plays a pivotal role by offering a clean, modular, and scalable solution to reduce greenhouse gas emissions from the electricity sector [6].

At the same time, the rapid deployment of PV systems introduces new operational challenges. The variability and intermittency of solar generation, determined by local meteorological conditions, create uncertainties that directly affect grid stability, market participation, and investment planning. As the share of renewables in the energy mix increases, the ability to anticipate production profiles becomes a cornerstone of system operation and long-term decarbonisation strategies. Accurate forecasts allow system operators to balance supply and demand, improve the scheduling of flexible resources such as storage, and reduce reliance on fossil-based backup generation. This is why forecasting is increasingly recognised as both a technical necessity and a strategic enabler for achieving climate-neutrality goals [5, 7, 8].

Within this context, research has focused on developing predictive models capable of translating the complex interaction between weather conditions, plant characteristics, and historical performance into reliable estimates of PV power output [9, 10]. A predictive model for PV power generation is based on the relationship between a set of input variables and an output value representing the electrical power or energy produced at a given time or over a defined time horizon. Typical input parameters include meteorological variables such as direct and diffuse solar irradiance, ambient and module temperature, wind speed and direction, relative humidity, and cloud cover, along with plant-specific information such as the orientation and tilt of PV modules, inverter efficiency, and historical power output. These models generally provide short to medium-term forecasts ranging from minutes to days of expected energy production. Model performance is assessed by comparing predicted values with actual measurements, thus evaluating both its accuracy in reproducing past behaviours and its generalisation capability in operational scenarios.

Accurate modelling and forecasting of PV production are essential for identifying potential malfunctions, optimising control strategies, and improving energy management and market participation [11–13]. However, the inherently variable and nonlinear nature of meteorological conditions makes it challenging to capture all the dynamics influencing power generation. Traditional physical and statistical models, though useful for describing general trends, often struggle to represent the temporal and spatial variability of weather-dependent variables [14, 15]. In this context, data-driven approaches and machine learning (ML) techniques have emerged as powerful tools capable of automatically learning complex relationships from data and providing accurate forecasts even under nonlinear and stochastic conditions [9].

The ability of a model to manage this transition effectively is crucial for its practical implementation in both energy generation and storage management [8].

In parallel with the increasing penetration of renewables, the demand for energy storage systems (ESS) has grown substantially. Storage technologies are essential to mitigate the variability of PV and wind generation, ensuring supply continuity, grid stability, and energy independence. Although lithium-ion batteries currently dominate the market due to their efficiency, energy density, and technological maturity [16–18], research efforts have expanded toward alternative technologies aimed at improving sustainability, safety, and cost-effectiveness. Among these, sodium-ion batteries are gaining attention for stationary applications thanks to their material abundance and lower costs; redox flow batteries,

particularly vanadium-based (VRFB), offer high safety, long lifetime, and scalability; while emerging technologies such as sulfur-based, metal-air, and solid-state batteries are being explored for their potential in next-generation energy storage. According to the International Energy Agency (IEA), the total installed battery capacity in the energy sector exceeded 2,400 GWh in 2023, representing a fourfold increase compared to 2020 and reflecting a diversification beyond lithium-based systems [19].

The integration of accurate PV forecasting with advanced energy storage solutions is expected to play a decisive role in the next phase of the energy transition [9, 10]. Forecasting enables operators to anticipate production profiles, optimise charging and discharging schedules, reduce curtailment, and limit reliance on fossil-based backup generation. Conversely, energy storage mitigates the uncertainty and variability of renewable generation, making forecasts more actionable for grid operators and energy markets. The synergy between forecasting and storage thus represents a crucial step toward higher renewable penetration, improved energy security, and enhanced system resilience [7, 8].

Within this framework, this thesis makes two distinct yet complementary contributions to existing literature. Firstly, it presents an advanced photovoltaic power forecasting approach. This approach is based on a systematic analysis of input variables and the construction of synthetic features. The aim is to enhance predictive accuracy under variable meteorological conditions. Secondly, it introduces an optimisation framework for battery energy storage systems with a particular focus on optimising the size of a vanadium redox flow battery according to economic performance criteria and technology-specific characteristics. Additionally, the proposed methodology facilitates consistent techno-economic comparisons between vanadium-based and lithium-ion storage systems, evaluating their competitiveness in terms of cost and suitability for photovoltaic integration. Overall, the thesis offers a unified methodological perspective on forecasting and storage optimisation, shedding light on the role of advanced battery technologies in facilitating the large-scale deployment of photovoltaic generation.

1.2 An overview on the PV power forecasting

1.2.1 Classification by Forecast Horizon

Photovoltaic power forecasting relies on methods that predict the power generation of a PV plant at different time intervals. The forecast horizon is the time period into the future for which the prediction is made. It has a significant impact on both the achievable accuracy and the type of model that should be used. Depending on the timescale, PV forecasting can serve a variety of purposes, ranging from real-time system control to long-term energy planning and investment analysis. PV power forecasting is typically categorised based on the forecast horizon and the type of forecasting model employed to estimate energy production. Regarding the forecast horizon, the literature identifies several categories [20–22]. These categories differ in terms of their temporal span, objectives, and the type of data and models required. Generally, seven forecasting horizons can be distinguished: very short-term (or ultra-short-term),intra-hour (nowcasting), intra-day, short-term, day-ahead, medium-term, and long-term forecasting.

- **Very short-term or ultra-short-term forecasting** refers to prediction intervals shorter than approximately 30 minutes. This horizon is essential for real-time monitoring and control of PV plants, being particularly relevant in smart grids and distributed generation systems. Such forecasts are employed for real-time electricity dispatch, grid balancing, short-term electricity price determination, and PV storage control.
- **Intra-hour forecasting**, also known as nowcasting, generally covers horizons from a few seconds up to one hour ahead. Nowcasting plays a critical role in guaranteeing the stability and quality of the power network, allowing operators to anticipate rapid variations in solar irradiance caused by transient clouds or atmospheric events. These forecasts are widely used to schedule operations on distribution systems and to enable short-term optimization of grid performance [23].
- **Intra-day forecasting** usually ranges from one to six hours and focuses on anticipating the PV generation profile over the course of a day. This type of forecasting allows operators to monitor specific electrical loads in local areas, plan distribution operations, and manage energy storage systems efficiently. Intra-day forecasts provide a balance between short-term accuracy and operational usefulness, supporting ap-

plications such as demand response, short-term grid scheduling, and hybrid energy management.

- **Short-term forecasting** generally spans from one hour to several days, and in some definitions up to one week. This type of forecasting is crucial for daily operational planning and for the participation of PV producers in electricity markets [24]. Accurate short-term forecasts allow energy suppliers to optimize unit commitment, dispatch strategies, and bidding operations, ensuring a reliable match between production and consumption. These forecasts typically integrate meteorological predictions obtained from numerical weather prediction (NWP) models with statistical or hybrid learning methods capable of correcting systematic biases.
- **Day-ahead forecasting** usually covers a period between 6 and 48 hours. Day-ahead forecasts are essential in electricity market operations, as they guide production scheduling, energy storage dispatching, and grid stabilization. Energy producers use these forecasts to submit bids in day-ahead markets and to coordinate generation with transmission and distribution planning. Because of their temporal scale, day-ahead forecasts often combine physical and statistical approaches, merging the spatially distributed information of NWP models with data-driven error correction methods.
- **Medium-term forecasting** extends from one week up to approximately one month. This time scale is mainly used for system operation planning, maintenance scheduling, and the management of energy resources across longer periods. For example, medium-term forecasts support the coordination of hybrid power plants, the maintenance of transformers and inverters, and the optimization of grid operation during seasonal transitions. These forecasts typically rely on physical and meteorological models, which describe the evolution of weather systems over several days, often complemented by statistical post-processing to enhance accuracy.
- **Long-term forecasting** encompasses horizons ranging from one month to one year or even longer. This category is primarily applied to strategic and economic planning, including investment assessment, capacity expansion, and long-term operation of power plants. Long-term forecasts also consider seasonal and interannual variability in solar resources, providing valuable information for the sizing of storage systems and for the evaluation of PV plant performance degradation. However, due to the

increasing influence of stochastic and climatic variability, the accuracy of long-term forecasts is typically lower than that of short-term predictions.

The classification of PV power forecasting by forecast horizon reflects the wide range of temporal scales and objectives involved in managing solar energy systems. Each horizon requires different levels of data resolution, computational complexity, and model sophistication. Shorter horizons demand real-time data and high-frequency modeling techniques, while longer horizons depend more on meteorological, climatological, and economic factors. Understanding these distinctions is fundamental for selecting appropriate forecasting strategies that ensure reliable, efficient integration of photovoltaic energy into modern power systems.

Having established the classification of PV power forecasting according to the temporal horizon, the subsequent analysis will focus on the principal modeling approaches employed in this field. Specifically, the following sections will examine physical, statistical, machine learning, and hybrid models, discussing their theoretical underpinnings, methodological features, and relative advantages and limitations with respect to different forecasting horizons.

1.3 Physical Models for Photovoltaic Power Forecasting

Physical, or physics-based, models aim to reproduce the underlying physical processes governing PV power generation. Unlike purely statistical or machine learning approaches, these models explicitly describe the energy conversion chain of a PV system, starting from meteorological and astronomical variables and leading to the electrical output of the plant. Their main strength lies in their interpretability and physical consistency, as they rely on first-principles formulations such as the laws of solar radiation, thermal balance, and the electrical behaviour of PV modules and inverters [25].

The modelling chain typically begins with the estimation of solar irradiance using numerical weather prediction outputs or satellite observations. From these data, the global, direct, and diffuse components of irradiance (GHI, DNI, and DHI) are derived and then transposed to the plane of array according to the system's tilt and azimuth. Transposition models account for the effects of module orientation, shading, reflection, and ground albedo, which can

significantly influence the effective irradiance incident on the modules [26, 27]. The next step involves modelling the cell temperature, as it has a strong impact on module efficiency. Common approaches include the Nominal Operating Cell Temperature model or more detailed thermal models that incorporate convective and radiative heat exchanges, ambient temperature, and wind speed [28].

Once irradiance and temperature are known, the electrical performance of the module is determined using an electrical model, often a single or two-diode equivalent circuit or simplified efficiency-based formulations. Downstream components such as inverters, cables, and balance-of-system elements are also modelled to account for conversion losses, mismatch effects, soiling, reflection, and wiring losses [29–31]. More advanced models include spectral and angle-of-incidence corrections to capture the impact of non-ideal optical conditions on energy conversion efficiency.

The main advantage of physical models is their interpretability: each stage of the modelling process can be analysed independently, enabling the identification of the subsystem that contributes most to forecasting errors. Physical models are also suitable when limited or no historical production data are available, as they do not rely on data-driven fitting, but rather on measurable physical parameters. Furthermore, they can incorporate detailed, site-specific information, such as module technology, orientation and local microclimate, making them adaptable to varying environmental conditions. However, their accuracy is highly dependent on the quality of the input data. Errors in irradiance forecasts, uncertain system parameters or simplified loss models can substantially reduce performance. Physical models are also more computationally demanding and may struggle to capture highly nonlinear or stochastic phenomena, such as fast-changing cloud dynamics, unless they are combined with data-driven methods in hybrid configurations [32].

Depending on their spatial and temporal resolution, physical models can be categorised into global, mesoscale, and microscale levels. Global models typically rely on coarse-resolution NWP data and are suited for day-ahead or longer-term forecasts. Mesoscale models, with finer spatial resolution, better capture regional meteorological features and local cloud systems, while microscale models apply local corrections accounting for terrain, shading, and microclimatic effects, often supported by satellite or sky-imager data [32].

Several recent studies have analysed and compared different physical modelling approaches. In particular, [32] conducted a comprehensive comparison of multiple modelling chains,

highlighting that irradiance decomposition and the transposition step represent the main sources of uncertainty, whereas inverter modelling has a relatively minor impact on overall accuracy. Similarly, [33] compared physical models applied to monocrystalline and polycrystalline modules using equivalent circuit representations with three to five parameters and different thermal models, such as NOCT and the Sandia model, showing that the choice of the temperature model has a decisive influence on forecasting performance. Finally, [34] compared purely physical and hybrid methods for day-ahead PV power forecasting, finding that the integration with statistical techniques often improves robustness and accuracy.

Although physical models demand precise and comprehensive input data, they offer a clear and insightful framework for comprehending the factors that affect solar energy production. Additionally, they establish a significant basis for hybrid approaches, where physical insights and data-driven techniques are merged to leverage the respective strengths of each domain.

1.4 Statistical Models

Statistical models represent one of the earliest and most fundamental approaches to photovoltaic power forecasting. They rely on the analysis of historical data to identify temporal and spatial dependencies between the measured PV output and explanatory variables such as irradiance, temperature, and meteorological conditions. Unlike physical models, which attempt to replicate the underlying energy conversion mechanisms, statistical methods focus on learning recurrent patterns and correlations directly from data, assuming that future system behavior will follow similar trends to those observed in the past. These models are particularly effective for short-term forecasting horizons, where recent observations provide a strong indication of near-future performance [20, 35].

Among traditional statistical techniques, time-series models such as the Autoregressive (AR), Autoregressive Moving Average (ARMA), and Autoregressive Integrated Moving Average (ARIMA) frameworks have been widely applied to solar power forecasting. These models describe the temporal evolution of PV output as a combination of autoregressive and moving-average components, effectively capturing persistence and seasonality in the data. To better handle daily and seasonal patterns, ARIMA is often extended to its Seasonal form (SARIMA), which incorporates periodicity into the model structure [36]. Regression-based methods, including Multiple Linear Regression (MLR) and Polynomial Regression,

have also been extensively used to correlate PV power with explanatory variables such as global horizontal irradiance, ambient temperature, wind speed, and humidity [37]. These approaches are simple, computationally efficient, and interpretable, which makes them suitable for operational forecasting and model validation.

However, purely statistical models are generally limited by their linear nature and by their dependence on the assumption of stationarity. They often fail to accurately represent nonlinear relationships and abrupt variations in PV output caused by fast-moving clouds, shading, or sudden irradiance fluctuations [35]. To partially overcome these limitations, several hybrid extensions have been proposed, such as integrating autoregressive models with exogenous variables (ARX, SARIMAX) or applying statistical decompositions like Principal Component Analysis to reduce dimensionality before regression [38]. In some cases, statistical approaches are also combined with physical features (e.g., clear-sky indices, NWP data) to improve robustness and interpretability [20].

Another aspect concerns the validation strategy. Since statistical models are sensitive to temporal dependence, cross-validation schemes designed for time series, such as rolling-origin or walk-forward validation, are typically employed to ensure that the model evaluation is consistent with its operational use [39]. These strategies make it possible to assess not only the point accuracy but also the stability of the models across different seasonal and meteorological conditions.

Beyond point forecasts, statistical methods have also been extended to probabilistic forecasting frameworks, where the objective is to provide prediction intervals or full predictive distributions instead of single values. Techniques such as quantile regression or distributional regression have been employed to quantify uncertainty in PV output, offering more informative inputs for decision-making in grid management and storage scheduling [40]. This shift reflects the growing recognition that uncertainty quantification is as important as point accuracy in operational forecasting.

In addition, multivariate and spatio-temporal extensions of statistical models have been developed to exploit the correlations between geographically distributed PV sites. Vector Autoregression (VAR) and its variants with exogenous variables (VARX) can capture both temporal dynamics and spatial dependencies, improving forecast accuracy at regional or aggregated levels [41]. These models are particularly useful for transmission system operators, who must manage large-scale PV integration across different locations.

In summary, statistical models provide a transparent, mathematically rigorous framework for PV power forecasting. Their simplicity, low computational demand and strong performance over short time periods make them a valuable component of forecasting pipelines, often serving as benchmarks or baseline models. However, their inability to capture nonlinear and rapidly changing dynamics has led to the development of hybrid and machine learning approaches. Currently, statistical models play an important role as baselines, calibration tools, and components within more advanced forecasting architectures.

1.5 Machine Learning and Data-Driven Models

Over the past decade, the rapid growth of computational power and the availability of large, high-resolution datasets have led to a paradigm shift in photovoltaic power forecasting, from traditional statistical methods toward machine learning and data-driven approaches. These models aim to automatically learn complex relationships and hidden patterns between input variables (such as irradiance, temperature, humidity, wind speed, and sky images) and PV output without requiring explicit physical formulations. Their flexibility, adaptability, and superior nonlinear modeling capabilities have made them particularly suitable for short-term and intra-hour forecasting applications [42, 43].

Among the most widely used ML models, Artificial Neural Networks (ANNs) have been extensively applied to PV forecasting due to their ability to approximate arbitrary nonlinear functions. ANNs consist of interconnected layers of neurons that iteratively learn from data by adjusting internal weights through optimization algorithms such as backpropagation. Feed-forward multilayer perceptrons (MLP) have been shown to outperform classical statistical models in short-term forecasts, especially when trained with a sufficient volume of representative data [44]. However, their performance can degrade under unseen weather conditions, and their "black-box" nature often limits interpretability.

To improve generalization, Support Vector Machines have been adopted as a powerful alternative for nonlinear regression [45]. SVMs rely on kernel functions (e.g., radial basis, polynomial) to project data into higher-dimensional spaces where linear regression can be applied more effectively. Their robustness against overfitting and capacity to handle small and noisy datasets make them suitable for environments with limited data availability [46]. Other learning algorithms, such as k-Nearest Neighbors (k-NN) or tree-based methods like

Random Forests (RF) and Gradient Boosting Machines (GBM), have also been applied to capture nonlinearities and complex feature interactions [41, 47].

With the emergence of deep learning, more advanced architectures have been proposed to model temporal and spatial dependencies in PV power generation. Convolutional Neural Networks (CNNs) are able to extract spatial features from sky imagery, satellite maps, or irradiance fields, making them ideal for image-based solar forecasting [38, 48, 49]. Recurrent Neural Networks (RNNs) and their advanced variants such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) excel at capturing sequential dependencies and temporal correlations in time-series data. LSTMs have demonstrated significant improvements in accuracy for both short-term and day-ahead PV forecasting, outperforming shallow networks and classical regression models [50, 51]. Hybrid deep learning frameworks combining CNN and LSTM layers have further enhanced performance by jointly learning spatial and temporal representations of solar dynamics. More recently, transformer-based architectures and graph neural networks have been investigated for solar forecasting, leveraging attention mechanisms and spatial correlations across PV sites [52].

Another important dimension concerns the quality and availability of input data. Real-world PV datasets often suffer from missing values, measurement errors, or sensor drifts, which strongly affect ML performance. Consequently, data preprocessing and cleaning strategies, such as outlier detection, gap filling, and normalization, are crucial steps in operational forecasting frameworks [41, 53]. Beyond raw data, feature engineering remains essential: lagged values, temporal indicators (hour of day, day of year), and clear-sky indices are widely used to enrich model input [54].

Recent research has also explored data fusion and feature engineering techniques to integrate multiple data sources, including NWP, sky cameras, and on-site sensors, to improve model robustness and generalization [55]. Feature selection methods such as Principal Component Analysis (PCA), Mutual Information (MI), or Recursive Feature Elimination (RFE) are frequently employed to identify the most relevant predictors and prevent overfitting. In addition, hyperparameter optimization algorithms, including Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Bayesian optimization, are often coupled with ML models to fine-tune their parameters and maximize performance [56].

Despite their strong predictive capabilities, ML models present several challenges. They require large and representative datasets, and their performance can deteriorate under

rare or extreme meteorological conditions. Moreover, they often operate as "black-box" systems with limited interpretability, which can be problematic for decision-making in grid management and system design. To address these issues, recent studies have investigated the integration of explainable artificial intelligence (XAI) techniques to interpret feature importance and model behavior [57]. Similarly, probabilistic ML approaches such as quantile regression forests, deep ensembles, or Bayesian neural networks have been introduced to explicitly quantify predictive uncertainty, which is increasingly required in market participation and risk-aware grid operation [40].

To provide a synthetic comparison, Table 1.1 summarizes the main families of ML and deep learning models applied to PV forecasting, highlighting their strengths and limitations in relation to different contexts. It should be emphasized, however, that the performance of these methods depends not only on the choice of algorithm, but also on the quality of the data and the preprocessing strategies applied before model training.

In summary, machine learning and data-driven models have revolutionized PV power forecasting by providing flexible, adaptive, and high-accuracy predictive frameworks.

1.6 Hybrid and Ensemble Models

Hybrid forecasting models have become a prominent approach in photovoltaic power prediction because they integrate diverse modeling paradigms (physical, statistical, and machine learning) to exploit their complementary advantages. The rationale is that no single modeling class can maintain high accuracy across all operating and meteorological conditions. Physical models retain interpretability and are grounded in solar conversion physics, but their performance depends heavily on the fidelity of numerical weather predictions. In contrast, statistical and ML models capture nonlinearities and hidden patterns, although they often suffer from limited interpretability and may generalize poorly outside their training domain. Hybrid frameworks thus seek to merge data-driven learning with physically informed constraints or to ensemble multiple predictive modalities, aiming to enhance robustness and generalisation [20, 22, 58].

Several categories of hybridization can be distinguished. In physics–data-driven hybrids, physical variables such as clear-sky indices or transposed irradiance are used as inputs or constraints for machine learning models, providing physical consistency while allowing

Table 1.1 – Summary of machine learning and deep learning models for PV power forecasting.

Model family	Strengths	Limitations
ANN / MLP	Flexible nonlinear modeling; good short-term accuracy	Require large datasets; poor interpretability
SVM	Robust with small/noisy datasets; kernel-based flexibility	Limited scalability; sensitive to kernel choice
Tree-based (RF, GBM, XGBoost)	Handle mixed data; capture complex interactions; good baseline	May overfit without tuning; less suited for temporal dependencies
k-NN	Simple; interpretable; no training phase	Computationally expensive at prediction time; weak with high-dimensional data
CNN	Excellent for spatial/image data (sky, satellite)	Require large annotated datasets; limited temporal modeling
RNN / LSTM / GRU	Capture sequential dependencies; state-of-the-art in time-series	High training cost; risk of vanishing gradients (RNN)
Hybrid CNN–LSTM	Joint spatial-temporal learning; strong empirical performance	Increased complexity; risk of overfitting
Transformers / GNN	Capture long-range dependencies and spatial correlations	Require large datasets; still emerging in PV forecasting
Probabilistic ML (QR, ensembles, BNN)	Provide uncertainty quantification	More complex evaluation and calibration

nonlinear learning [34]. In statistical–ML hybrids, autoregressive or ARIMA components are often combined with ANNs, SVMs, or boosting methods to capture both linear and nonlinear dependencies. Multi-stage pipelines also exist, where one model is used to correct systematic errors from another, or where decomposition techniques are employed to separate the signal into more regular components before prediction.

One common hybrid approach is to pair an ANN with a GA, which optimises network weights and hyperparameters globally to avoid local minima issues. In [59, 60] it is shown that ANN+GA configurations can enhance PV forecasting performance compared to standalone neural networks. More recent studies confirm that GA-based hyperparameter tuning for FF-ANN architectures improves accuracy in operational forecasting tasks [61]. Other metaheuristic optimizers such as Particle Swarm Optimization or Differential Evolution have also been successfully coupled with ML predictors, further highlighting the role of global search strategies in hybrid frameworks.

Another hybrid methodology involves decomposition–prediction pipelines: the irradiance or power time series is first decomposed (e.g. via wavelets, empirical mode decomposition, or variational mode decomposition), then each component is forecasted with a tailored predictor (e.g. LSTM, SVM, or Random Forest), and the results are recombined for final output. These frameworks are valuable in managing non-stationarity and multiscale dynamics in PV signals, especially under rapid variability [62]. Studies have shown that decomposition can isolate high-frequency components linked to cloud transients, which can then be addressed with specialized models better suited to capturing fast fluctuations.

Ensemble and hybrid-ensemble strategies have also gained traction: multiple predictors run in parallel and their outputs are combined (e.g. by averaging, weighted voting, stacking, or using a meta-learner). In [63], an ensemble based on weather classification is proposed and demonstrated to outperform individual models in PV sites across Germany. A comprehensive review of ensemble forecasting methods in solar applications [64] confirms that ensemble and hybrid techniques are increasingly considered state of the art for operational PV forecasting. In addition, dynamic hybrid ensembles that select the most suitable base predictor per instance have shown robustness gains, as demonstrated in solar plant case studies [65]. Recent works also emphasize the use of ensemble approaches to provide probabilistic forecasts, where the diversity of models is exploited not only to improve mean accuracy but also to characterize forecast uncertainty [40].

From an operational perspective, hybrid and ensemble models are particularly valuable for short-term and day-ahead horizons, where the interaction between fast cloud dynamics and meteorological forecast errors makes the prediction problem highly challenging. For intraday applications, decomposition-based hybrids are often preferred to handle rapid ramps, while for day-ahead horizons, ensembles combining NWP-based physical models with statistical or ML corrections have proven effective in reducing systematic bias [22].

Despite these advantages, hybrid models come with challenges, including greater computational burden, hyperparameter tuning complexity, and risk of overfitting due to increased structural complexity. Their success also depends critically on high-quality, representative training data and on the careful design of validation schemes to avoid overestimating performance. Moreover, while ensembles generally improve accuracy, they may reduce interpretability, which is a crucial consideration for system operators and market participants. Nonetheless, by combining the interpretability of deterministic models with the flexibility of data-driven techniques, hybrid and ensemble approaches remain among the most promising frameworks for accurate PV power forecasting across temporal horizons.

1.7 The Nature and Preprocessing of Data for PV Power Forecasting

The development of a photovoltaic power forecasting model is a complex and multi-layered process, in which the quality and representativeness of input data play a decisive role. Regardless of whether the approach is based on physical principles, statistical techniques, or machine learning methods, predictive performance ultimately depends on the availability of reliable data and the care with which such data are treated in the preliminary phases preceding model training. In general, a forecasting model operates by learning the relationship between a set of inputs (features), typically meteorological and operational variables such as solar irradiance, ambient temperature, wind speed, humidity, or derived indices, and an output (label), which corresponds to the predicted photovoltaic power generation at a given instant or over a defined forecasting horizon. The clear definition of this input–output structure constitutes a fundamental prerequisite for understanding the subsequent stages of model development.

Among meteorological variables (inputs), solar irradiance plays the most critical role, as it constitutes the direct driver of the photovoltaic conversion process. Other factors such as ambient temperature, wind speed, humidity, and cloud cover also influence generation, although generally to a lesser extent. Consequently, access to accurate irradiance measurements is essential to reduce forecasting uncertainty, especially in short-term horizons. To improve robustness, irradiance is often transformed into derived indices such as the clear-sky index or clearness index, which normalize observed values with respect to expected clear-sky conditions and reduce nonstationarity in historical series [20].

Once collected, raw data must be subjected to a rigorous preprocessing pipeline to ensure their reliability and suitability for use in ML applications. Gaps due to sensor failures or communication issues can be addressed through imputation techniques such as forward/backward filling, spline interpolation, or model-based approaches like regression or k-nearest neighbors [66]. Outliers, on the other hand, are commonly identified and removed using statistical methods (Z-score, interquartile range) or robust regression [67]. Subsequently, normalization or standardization procedures are applied to harmonize predictors, while domain-specific transformations prove particularly useful for highlighting the influence of cloud cover on irradiance dynamics [20].

Alongside preprocessing, feature engineering represents a crucial step to enhance the explanatory power of datasets. In addition to raw meteorological variables, it is common to construct derived variables capable of capturing temporal dependencies and cyclic patterns. Lagged values of irradiance and power provide autoregressive information; moving averages and rolling statistics make it possible to capture short-term fluctuations, while temporal indicators such as hour of the day, day of the year, or seasonal terms make cyclical recurrences explicit. More sophisticated approaches involve the creation of synthetic variables obtained through the transformation or combination of existing measurements. For example, [68] proposed an LSTM model that integrates synthetic weather forecasts with historical irradiance data and sky-type classifications, achieving a 44.6% improvement in daily forecast accuracy. Similarly, [69] applied PCA to identify and combine twelve key meteorological variables, obtaining superior performance compared to baseline models. Other studies confirm the benefits of feature engineering in integrating heterogeneous sources: [54] used sky cover indices derived from total sky imagers, infrared radiation, and GHI measurements as inputs to an artificial neural network, while [70] combined lagged power values, moving averages, and time-based variables in a stacked ensemble

model. More recently, CNNs have been employed to automate feature extraction from sky images, showing great potential both for solar radiation forecasting and for fault detection in photovoltaic systems [71].

When working with large and heterogeneous datasets, redundancy and collinearity among predictors may reduce efficiency and increase the risk of overfitting. For this reason, dimensionality reduction and feature selection techniques such as PCA, mutual information, or recursive feature elimination are frequently adopted to retain only the most relevant variables [55]. Another crucial aspect concerns the division of data into training, validation, and test sets. Unlike other ML applications, random partitioning is not suitable for time-series forecasting, as it risks introducing information leakage and distorting evaluation. Instead, time-aware validation strategies such as rolling-origin (walk-forward) validation or blocked cross-validation are recommended to preserve chronological order and better reproduce operational conditions [39]. In addition, preprocessing may include detrending procedures aimed at removing long-term effects, such as system degradation, allowing models to focus on short-term variability that is more relevant for operational horizons.

Ultimately, the forecasting model is trained to learn the mapping between inputs and outputs, its hyperparameters are optimized on the validation set through techniques such as grid search, random search, or evolutionary algorithms, and its generalization capacity is evaluated on an independent test set. Performance is usually quantified using deterministic error metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE); however, in operational contexts where risk management plays a crucial role, probabilistic approaches and calibration metrics are increasingly being adopted [20, 41].

1.8 Battery Energy Storage Systems: Technologies and Role in Renewable Integration

Accurate forecasting of photovoltaic generation, as discussed in the previous sections, is a fundamental step to reduce operational uncertainty and to achieve a better alignment between generation and demand. However, for forecasts to become truly useful for the power system, it is necessary to rely on flexibility resources capable of absorbing the deviations between expected profiles and the actual behaviour of generation and loads. In

this context, Battery Energy Storage Systems (BESS) represent one of the most versatile and technologically mature solutions. They enable the translation of forecasting information into concrete energy management actions, providing a wide range of services [72, 73].

Among the main services enabled by BESS it is possible to distinguish:

- Time-shifting of generation: storing energy during periods of maximum PV production and releasing it during hours of higher demand allows the maximisation of self-consumption and reduces reliance on conventional generation. Accurate forecasts enable the planning of charge and discharge cycles in advance, minimising renewable energy losses due to curtailment [74].
- Peak shaving: BESS can discharge energy during peak demand periods of the grid or local loads, reducing the power drawn from the network and alleviating stress on electrical infrastructure and transformers. This provides a dual benefit: on the one hand, the reduction of congestion and grid reinforcement costs, and on the other, the containment of electricity tariffs for end-users. Integration with PV forecasting allows discharges to be scheduled precisely when the predicted load exceeds specific thresholds, minimising operational costs[75].
- Frequency regulation and fast reserves: thanks to their near-instantaneous response times, BESS can rapidly compensate imbalances between supply and demand, thereby stabilising grid frequency. Forecasts of renewable generation support the definition of the reserve margin that can be allocated to such services without compromising availability for other functions [76].
- Voltage support and power quality: storage systems can contribute to voltage control and to the compensation of reactive power fluctuations, improving the reliability of local networks. Forecasts of PV injection, combined with load models, make it possible to identify critical periods in advance and to optimise BESS operation for system stability [77].
- Curtailment reduction and renewable penetration increase: during hours when PV production exceeds the absorption capacity of the grid, BESS can store the energy that would otherwise be curtailed, ensuring a more efficient use of available renewable resources [78].

The effectiveness with which these services are delivered depends jointly on the storage technology and the Battery Management System, which plays a crucial role in ensuring operational safety, optimising charge–discharge cycles, and mitigating degradation mechanisms. In a framework where PV generation can be forecasted with increasing accuracy, BESS become the instrument that allows such information to be fully exploited: the ability to schedule charge and discharge set-points in advance, based on forecasts and grid or market signals, enables the maximisation of both economic and systemic benefits associated with renewable integration.

A wide range of electrochemical storage technologies has been developed and deployed in stationary applications, each with specific advantages, limitations, and levels of maturity. Among the most established are lead–acid batteries, widely used thanks to their low cost and a well-developed recycling chain, but characterised by low energy density and limited cycle life, which restrict their use to small-scale or backup systems [79]. Lithium-ion batteries currently dominate the stationary storage market, driven by high efficiency, fast response times, and continuous cost reductions supported by economies of scale in the automotive sector [80, 81].

Sodium–sulfur batteries represent another commercially mature option, offering high energy density and suitability for long-duration applications such as daily load shifting [82]. Their operation at elevated temperatures, however, introduces technical and safety challenges that limit widespread adoption. Redox flow batteries (RFB), particularly VRFB, are emerging as a promising alternative for grid applications [83, 84].

Alongside these consolidated technologies, research is increasingly exploring alternative chemistries such as sodium-ion, zinc–bromine, and metal–air batteries. Although still at lower levels of technological maturity, they offer potential advantages in terms of sustainability, resource availability, and cost reduction, and may in the future complement or partially replace more established solutions [79].

Overall, the diversity of available storage technologies indicates that no single solution is suitable for all applications [85]. Lithium-ion batteries are best suited to short-duration and high-power services, while flow batteries and sodium–sulfur systems are more appropriate for long-duration storage. Lead–acid batteries, despite their limitations, remain a cost-effective option for small-scale and backup applications. The selection of the most appropriate technology must therefore be aligned with the temporal scale of the required

service and the specific operational context, while also accounting for cost, lifetime, and safety. Within this framework, the integration of accurate PV forecasting with properly sized and operated BESS emerges as a key enabler to maximise renewable utilisation and to ensure the economic sustainability of hybrid energy systems.

1.9 An overview on optimization models for the Battery Energy Storage System

In parallel with the technological progress of battery storage, the scientific literature has extensively investigated optimization models as a methodological tool to analyse the operation and the sizing of BESS. These models are characterized by a common conceptual structure consisting of input parameters, decision variables, technical and economic constraints, and an objective function. Input parameters typically include demand and generation profiles, electricity prices, battery efficiency and degradation characteristics. Decision variables represent the quantities to be optimized, such as charging and discharging schedules, the state of charge (SOC), or the storage capacity itself when investment decisions are included. Constraints enforce feasibility, ensuring that operational and market rules are respected. The objective function reflects the purpose of the model, which can vary from profit maximization through arbitrage and ancillary services, to cost minimization or multi-objective formulations that balance economic, technical, and environmental criteria [80].

The mathematical formulation of the optimization problem depends on the application. Linear programming (LP) and, more prominently, mixed-integer linear programming (MILP) are widely adopted, as they allow accurate representation of operational and logical constraints while remaining computationally tractable. These methods have been employed to optimize day-ahead scheduling of storage systems and to co-optimize arbitrage and frequency regulation, highlighting the value of simultaneous participation in multiple electricity markets [86]). When the representation of nonlinearities such as SOC-dependent efficiency or temperature dynamics is unavoidable, nonlinear programming (NLP) or mixed-integer nonlinear programming (MINLP) formulations are used, albeit at the cost of significantly higher computational complexity [87].

Beyond deterministic formulations, stochastic optimization has become increasingly relevant, as it allows the explicit representation of uncertainty in renewable generation and electricity prices. Scenario-based stochastic programming models, for example, capture a range of possible future realizations, enabling strategies that hedge against risk and improve system resilience. Robust optimization methods instead focus on guaranteeing feasibility under worst-case deviations, producing more conservative but reliable operating schedules. These approaches are especially important in contexts where forecast errors of photovoltaic and wind generation can severely impact the economic viability of storage operation.

For real-time control, dynamic programming and reinforcement learning approaches have been explored. Dynamic programming is theoretically well-suited to problems where future states depend on current actions, but suffers from the curse of dimensionality. Reinforcement learning, in contrast, learns control policies directly from interactions with an environment and has been applied to the real-time scheduling of batteries under uncertain conditions, including degradation and thermal constraints [88]. In parallel, metaheuristic methods such as genetic algorithms, particle swarm optimization, and simulated annealing have been employed to tackle large-scale or highly nonlinear problems. Although they do not guarantee global optimality, they can efficiently explore large solution spaces, making them suitable for long-term planning studies or for comparing different storage technologies [87].

Within this broader framework, specific modeling approaches have been developed for different battery technologies. For lithium-ion (Li-ion) batteries, optimization models typically emphasize charge and discharge scheduling while explicitly accounting for degradation [89]. The inclusion of degradation costs, either through marginal cost coefficients or through empirical cycle-aging models, has been shown to be decisive for the correct evaluation of long-term profitability [90]. Studies on grid-scale Li-ion systems often employ MILP or stochastic programming formulations to evaluate trade-offs between short-term arbitrage profits and long-term performance losses, or to co-optimize investment and operation in uncertain price environments [91].

For VRFBs optimization models must incorporate specific characteristics that differ from Li-ion systems. The decoupling between energy capacity (determined by electrolyte volume) and power capacity (determined by stack size) requires formulations that jointly optimize both design variables [92]. Moreover, efficiency depends on flow rate and operating power, while auxiliary losses from pumping play a non-negligible role. [93] proposed an MILP formulation with convex hull linearizations to represent variable efficiencies and capacity

fade, showing that simplified assumptions of constant efficiency can significantly overestimate revenues. Other studies [94] explicitly modeled flow management and rebalancing strategies, linking electrochemical dynamics with empirical estimates of auxiliary consumption. These examples demonstrate how technology-specific features must be carefully embedded in optimization models to provide realistic techno-economic assessments.

Overall, the literature illustrates a wide spectrum of optimization methodologies, ranging from exact mathematical formulations to heuristic and learning-based approaches. Each class of models is suited to specific contexts: deterministic MILP for day-ahead scheduling in market environments, stochastic formulations for uncertain renewable integration, nonlinear models for detailed device-level representations, and reinforcement learning for adaptive real-time operation. A unifying theme across these contributions is the recognition that optimization models are not only useful for short-term scheduling, but are essential tools for identifying the correct sizing and configuration of storage systems, particularly in hybrid energy systems with multiple renewable sources. By systematically accounting for operational constraints, technology-specific features, and uncertainty, these models enable the design of economically competitive and technically viable solutions that maximize both the value of storage and the integration of renewable energy.

The next section outlines the structure of the thesis, offering a brief overview of the content covered in each chapter.

1.10 Thesis Outline

This thesis is structured to address the main challenges related to photovoltaic production forecasting and its integration with storage systems in a progressive manner, moving from methodological foundations to optimisation applications. Each chapter is organized to build upon the results of the previous one, forming a coherent framework that addresses both forecasting accuracy and techno-economic implications of storage deployment.

Chapter 2 focuses on the methodological foundations of photovoltaic power forecasting. The chapter introduces a data-driven forecasting framework based on historical observations, emphasizing the importance of data quality and preprocessing. Special attention is devoted to the identification and treatment of missing data, outliers, and measurement uncertainties, which represent critical issues for model reliability. The chapter also provides a comparative

assessment of classical statistical models, serving as baseline approaches against which more advanced methods can be evaluated. This first part lays the groundwork by clarifying how input data and methodological assumptions directly influence forecast performance.

Chapter 3 explores photovoltaic power forecasting using real weather forecast data and evaluates the use of synthetic variables in scenarios with limited data availability. Two models are developed and compared: the Base-model, which relies on commonly available meteorological variables and represents the standard operational approach, and the SF-model, in which some of these variables (irradiance and cloud cover) are replaced by synthetic indices designed to capture the nonlinear interactions among atmospheric variables. The analysis assesses whether synthetic variables can serve as reliable substitutes while maintaining good forecasting accuracy even under incomplete data conditions. Model performance is evaluated through RMSE and the coefficient of determination, with attention to the balance between model complexity, interpretability, and predictive skill.

Chapter 4 shifts the perspective from forecasting to system integration, proposing an optimization model for the sizing of BESS in PV applications. The model focuses on VRFB, a technology of increasing relevance due to its scalability, safety, and suitability for long-duration storage. The optimization framework considers both technical constraints and economic criteria, enabling the identification of optimal system configurations under different operating scenarios. In addition, a comparative techno-economic analysis with lithium-ion batteries is carried out, providing insights into the relative competitiveness of alternative storage technologies. The chapter thus bridges the forecasting framework developed in the previous chapters with practical applications in renewable energy integration and system planning.

Finally, Chapter 5 summarizes the main contributions of the thesis and discusses their broader implications. The conclusions highlight how advanced forecasting techniques, when coupled with optimized storage deployment, can significantly improve the operational efficiency and economic viability of PV systems. The chapter also outlines the limitations of the proposed approaches and suggests avenues for future research, including the integration of probabilistic forecasting, hybrid model ensembles, and advanced optimization methods for storage systems in multi-energy contexts. By doing so, the thesis not only addresses current challenges but also provides a perspective on how forecasting and storage will evolve in the next phase of the energy transition.

Chapter 2

PV power forecasting based on historical weather data

2.1 Motivation and Objectives

The increasing penetration of photovoltaic generation in modern power systems has made accurate power forecasting a central requirement for grid stability, market participation, and plant-level optimization. Unlike conventional generation, PV production is inherently variable and strongly dependent on meteorological conditions, thus introducing a high degree of uncertainty into operational planning and grid management. Consequently, the development of forecasting methodologies has received growing attention, ranging from traditional physical models based on irradiance and temperature equations to more flexible statistical and ML approaches capable of capturing nonlinear dependencies between meteorological variables and PV output [95–97].

Among the main ML approaches applied to PV power forecasting are linear and polynomial regression models [20], support vector machines (SVM) [98], ANN, and their advanced architectures, such as recurrent neural networks (RNN) and long short-term memory (LSTM) networks, which are particularly suitable for time-series prediction [10]. Further developments include ensemble methods, such as Random Forests, eXtreme Gradient Boosting (XGBoost), and LightGBM, known for their ability to handle complex datasets and nonlinear dependencies. Hybrid approaches that combine physical or statistical models

with ML algorithms have also been proposed to enhance the stability and generalisation of predictions [20, 99, 100].

Traditionally, ML models have been trained and validated using historical datasets that include measured meteorological and power data [98]. While this approach often yields good performance in controlled environments, several studies have shown that prediction accuracy may decrease significantly when the models are applied to real weather forecast data instead of historical data [101–103]. This observation underscores the importance of developing forecasting frameworks that directly leverage meteorological forecasts, thereby improving robustness and applicability in operational contexts where decision-making depends on predicted rather than observed variables [10, 104–106].

Among these, ensemble machine learning (EML) algorithms are among the most widely adopted. The term ensemble refers to the combination of multiple individual models into a single predictor with higher accuracy than its individual components. Ensemble approaches can be based on bagging (e.g., Random Forest), where models are trained in parallel on different subsets of the dataset, or on boosting (e.g., Gradient Boosting, AdaBoost, LightGBM, XGBoost, CatBoost), where weak learners are trained sequentially and each new model aims to correct the errors of its predecessor [107]. This mechanism makes boosting models particularly effective in complex forecasting tasks, as they progressively construct a strong predictor from simple and weak models.

Gradient Boosting, employed in this chapter, belongs to this class of algorithms. The choice of GB is motivated by its suitability for data-driven PV power forecasting problems in which the target variable is driven by nonlinear interactions among heterogeneous meteorological predictors. Compared to linear or kernel-based methods, GB can capture threshold effects and regime changes (e.g., low-irradiance vs clear-sky conditions) through tree-based partitioning, without requiring strong parametric assumptions. Moreover, boosting ensembles are typically robust when the input features are partially correlated and exhibit non-stationary statistical properties, as often occurs in meteorological time series over different seasons and operating regimes. From a practical perspective, GB (and its efficient implementation in LightGBM) offers an advantageous trade-off between predictive performance, computational cost, and interpretability. The model can be trained efficiently on large datasets with relatively limited feature engineering, and it provides a direct way to quantify the relevance of predictors (e.g., feature importance), which is useful to support the comparative analysis between local measurements and satellite-derived inputs. Nonethe-

less, GB also presents limitations that are particularly relevant for time-series applications. First, as a supervised regression model based on tabular samples, GB does not explicitly enforce temporal causality: temporal dynamics must be provided through the input features (e.g., timestamps, lagged variables, or calendar descriptors). Second, boosting methods can be sensitive to hyperparameter choices and may overfit if the ensemble complexity is not properly regularized (e.g., excessive depth or number of trees). Finally, when the training and validation sets are not separated in a temporally consistent way, information leakage may inflate performance estimates; for this reason, the splitting strategy adopted in this study is discussed explicitly in the following. GB is a sequential training process involving a set of M decision trees. The first tree is trained using the input features (X) and labels (y), while subsequent trees are trained on the residual errors of the previous ones. The final model thus consists of an ensemble of M trees that collaboratively improve predictions.

Figure 2.1 illustrates the architecture of a GB model. On the left-hand side, the initial datasets are represented, in the middle the sequence of decision trees trained iteratively, and on the right the combination process that produces the final predictor. Each tree does not operate independently but integrates corrective information from the errors of the previous models, ultimately converging into a more accurate overall predictor.

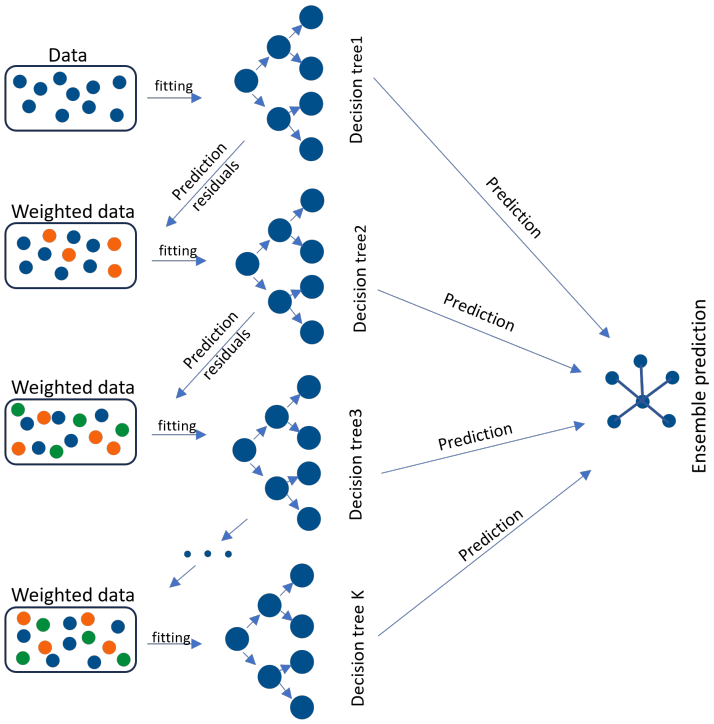


Figure 2.1 – Schematic representation of Gradient Boosting architecture: from input datasets, through sequential decision trees, to the final ensemble predictor.

The training objective is the minimization of a *loss function*, which measures the discrepancy between predicted and actual values. Several types of loss functions exist depending on the problem. In regression tasks, the most commonly used are the MSE, the MAE, and the Quantile Loss [108], the latter being particularly useful for quantile-based prediction intervals or risk-sensitive applications. At each iteration, the algorithm computes the gradient of the loss function with respect to the current predictions and trains a new weak learner to reduce this gradient. The process continues until a stopping criterion is met, such as a maximum number of trees or a minimum error threshold [109].

Building a predictive model through ML requires a data pre-processing phase. Along with meteorological parameters, the initial datasets include temporal information (timestamps) linking features (X) and labels (y). Since meteorological variables often differ in scale and dispersion, standardization is performed to rescale them to a common range with zero mean and unit standard deviation [110], thus facilitating training and improving predictive performance.

The training phase aims to define a general function $F(X)$ capable of estimating the PV power output from unseen input data. In supervised learning, datasets are typically split into three subsets:

- **Training set** (approximately 80% of the data), used to train the model.
- **Validation set**, used to tune hyperparameters and prevent overfitting.
- **Test set** (approximately 20% of the data), used to evaluate the final model on independent observations.

A random split with a fixed random seed ensures reproducibility and statistical comparability between subsets. The adopted data splitting strategy balances robust model selection and realistic performance evaluation. Although a purely temporal split is commonly used in time-series forecasting, it may bias hyperparameter tuning when the validation set covers only limited seasonal or meteorological regimes; therefore, a random split with a fixed seed is employed for training and validation to ensure representative sampling across heterogeneous conditions. To avoid optimistic performance estimates due to temporal correlation, the final evaluation is performed on an independent seven-day consecutive test set fully excluded from training and validation.

The performance of GB is highly dependent on its hyperparameters, such as the number of trees, the maximum depth of each tree, the number of leaves, the loss function, and the learning rate. Hyperparameter tuning can be carried out through deterministic or stochastic search strategies (e.g., grid search, random search), which explore different combinations within predefined ranges.

The final evaluation of the model is carried out by comparing predictions with actual test labels. The most common performance metrics in regression tasks are the Root Mean Square Error (RMSE) and the coefficient of determination (R^2), which respectively measure the average prediction error and the proportion of variance explained by the model.

Beyond the modeling technique, a fundamental determinant of forecasting accuracy is the nature of the meteorological input data. PV forecasts may rely on high-resolution in-situ measurements, obtained from local weather stations and pyranometers, which provide site-specific and highly correlated information but require dedicated instrumentation and are not always available. Alternatively, satellite-based data ensure wide coverage and accessibility at the expense of reduced spatial and temporal resolution. In [105], it is shown that the accuracy of LightGBM-based PV forecasts depends strongly on the quality and resolution of irradiance inputs. In [111], a deep neural network is trained using multi-temporal geostationary satellite imagery. In [112] satellite-derived irradiance products combined with numerical weather prediction are employed to estimate and forecast regional PV power without dense local measurements, underlining the scalability of this approach. In [113] geostationary satellite images are processed to directly forecast PV power (1–6 hours ahead) without an intermediate irradiance step, demonstrating operational benefits in utility-scale plants. In [114] the value of satellite-derived irradiance for gauging and verifying gridded forecasts is discussed, supporting their role as a practical alternative when ground sensors are sparse. More recently, [115] demonstrates that a hybrid LSTM-Gated Recurrent Unit (GRU) architecture trained exclusively on satellite-driven data enables reliable PV power forecasting in data-scarce regions, achieving competitive accuracy with ground-based approaches. This result indicates that advanced data-driven methods can partially compensate for the lower resolution of satellite products, making them a viable option in contexts where local meteorological measurements are unavailable. However, the literature still lacks systematic comparative analyses that evaluate, under the same methodological framework, the impact of local versus satellite data on forecasting accuracy.

This gap is particularly relevant from an operational perspective. On the one hand, local meteorological measurements are highly correlated with actual PV output and can maximize accuracy but are costly and available only to plant operators, limiting model generalizability. On the other hand, satellite data are broadly accessible and scalable across multiple sites, but their reduced resolution may compromise short-term accuracy. Quantifying this trade-off between accuracy and accessibility is therefore essential to guide both system operators and researchers in developing forecasting tools that are not only theoretically sound but also practically deployable at scale.

In this context, the objective of this chapter is to investigate the role of meteorological data sources in shaping PV power forecasting performance. To this end, two gradient boosting models are developed and compared: the first trained on historical weather data from local weather stations and pyranometers, and the second trained on equivalent meteorological features obtained from satellite-based services via application programming interfaces (API). Using real datasets from PV plants of different sizes and geographic locations, the analysis aims to:

(i) evaluate the predictive accuracy of both approaches through statistical error metrics,

(ii) assess the extent to which satellite data can serve as a practical alternative to local measurements, and

(iii) provide insights into the trade-off between model performance and data availability.

By addressing this research question, the chapter contributes to the broader debate on the development of data-driven forecasting systems for photovoltaics, highlighting the methodological and operational implications of input data selection.

2.2 Model Set Up

The accurate prediction of PV power requires not only high-quality input data but also an appropriate modeling framework. In this study, ML techniques were adopted for their ability to capture complex, non-linear relationships between meteorological variables and PV output. Among the various ML methods, GB was selected due to its proven effectiveness in handling heterogeneous datasets and providing robust predictions across different forecasting horizons. Furthermore, recent implementations such as the LGBM improve computational efficiency through optimized tree growth strategies, making them particularly

suitable for forecasting tasks in which input features exhibit complex dependencies, as is the case for PV power generation.

The dataset considered in this work consists of one year of PV production and meteorological data (September 2022 – September 2023) collected from three photovoltaic plants differing in size and geographic location across Italy. PV power values are normalized with respect to the rated peak capacity of each plant (kW/kWp), thereby enabling direct comparison across systems with different scales. The meteorological variables include solar irradiance, air temperature, relative humidity, wind speed, and wind direction, along with temporal information derived from timestamps to capture diurnal and seasonal patterns.

Since the meteorological variables differ in scale and dispersion, a standardization step was applied. Each feature was rescaled to have zero mean and unit variance, facilitating model convergence and ensuring a consistent scale across variables [110]. Outliers, especially in irradiance and wind speed, were retained to preserve the variability associated with rapidly changing weather conditions. The dataset was then divided into three subsets: 80% of the data was used for training, 20% for validation, and an independent test set was created by selecting seven consecutive days entirely excluded from both training and validation. This splitting strategy ensures a realistic evaluation of model generalization on temporally non-overlapping data. Random splitting with a fixed random seed was employed to guarantee reproducibility and statistical comparability.

Gradient boosting models were implemented using the LightGBM framework. GB constructs an ensemble of weak learners, typically decision trees, where each tree is trained to correct the residual errors of its predecessors. This sequential process enables the model to progressively focus on the most difficult patterns, ultimately improving predictive accuracy. Model performance is strongly influenced by hyperparameters such as the number of trees, maximum tree depth, learning rate, number of leaves, and regularization terms. A randomized search procedure (RandomizedSearchCV) was employed to explore the hyperparameter space efficiently, striking a balance between accuracy and computational cost. The selected configuration included a quantile loss function, 200 estimators, maximum depth of 20, and a learning rate of 0.05.

The final evaluation of the models was carried out on the independent test set. Two statistical metrics were employed: the Root Mean Squared Error (RMSE), which measures the average deviation between predicted and observed PV power values,

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - F(X_i))^2}\,, \tag{2.1}$$

and the coefficient of determination (R^2), which quantifies the proportion of variance in PV power explained by the model,

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - F(X_i))^2}{\sum_{i=1}^{n}(y_i - \bar{F}(X))^2}\,. \tag{2.2}$$

Finally, two different sources of meteorological input data were considered to assess their impact on forecasting performance. In Scenario A (SA), localized measurements from nearby weather stations and pyranometers provide high-resolution, site-specific data. In Scenario B (SB), the same set of meteorological variables is obtained from satellite-based services accessible through APIs. A detailed schematic and statistical comparison of these scenarios are presented in Section 2.3. This design enables a direct evaluation of the trade-off between accuracy and accessibility in PV forecasting applications.

2.3 Scenario

A schematic description of the analyzed scenarios is reported in Fig. 2.2. This work considers two different scenarios, which differ in the source of meteorological data used to train the machine learning model:

- **Scenario A (SA):** data are collected from weather stations located within a radius of 500 meters from the selected PV plants. In addition, the plants are equipped with on-site pyranometers that provide instantaneous solar irradiance measurements. The available variables include air temperature T (°C), relative humidity rh (%), wind speed ws (m/s), wind direction wd (degrees), and solar irradiance G (W/m^2).
- **Scenario B (SB):** in this case, weather data are retrieved online from [116]. The data are obtained from satellite surveys and made available through an API, starting from the geographic coordinates of each plant. The API provides, in addition to the

variables considered in SA, a much broader set of descriptors (e.g., cloud coverage, direct and global irradiance, precipitation probability, atmospheric pressure, etc.). However, to ensure a fair comparison with SA, only the variables common to both scenarios are included in the present analysis, while the additional descriptors have been intentionally excluded.

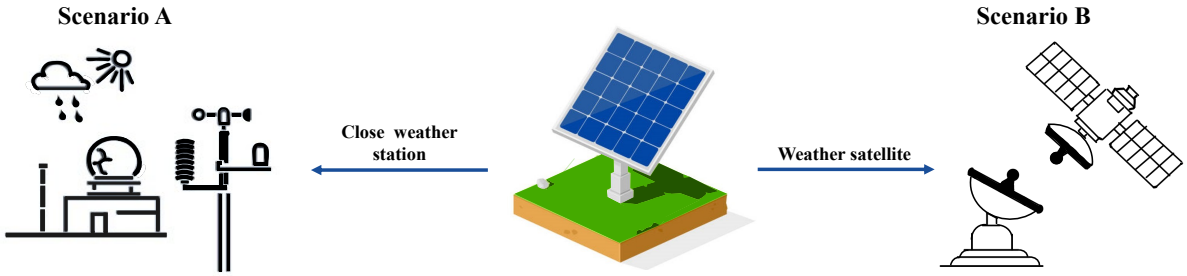


Figure 2.2 – Schematic representation of the two scenarios considered in this work. In Scenario A (SA), weather and irradiance data are collected from local weather stations and pyranometers placed within 500 m of the PV plants. In Scenario B (SB), weather data are retrieved from satellite-based services accessible via API, starting from the geographic coordinates of the plants.

The PV production and meteorological data analyzed span a one-year period, with a temporal resolution of 15 minutes. Three PV plants of different sizes and geographical locations in Italy are considered. This choice allows us to evaluate model performance under heterogeneous conditions: not only do the plants differ in their power profiles, but the different locations also enable us to assess the accuracy of API-based satellite data retrieved from distinct coordinates.

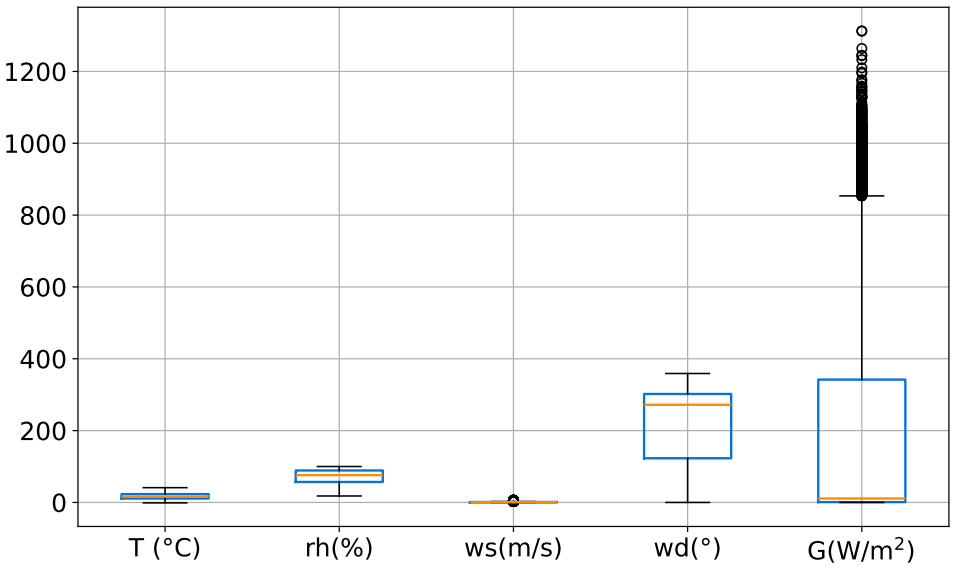
All PV power data are normalized with respect to the rated peak power of each plant (kWp) and expressed in kW/kWp, thus ensuring comparability across systems. Pre-processing steps were applied consistently to both scenarios. For clarity, Table 2.1 summarizes the variables shared by the two analyzed scenarios.

Table 2.1 – Example of the first five rows of the structure of the initial data frame.

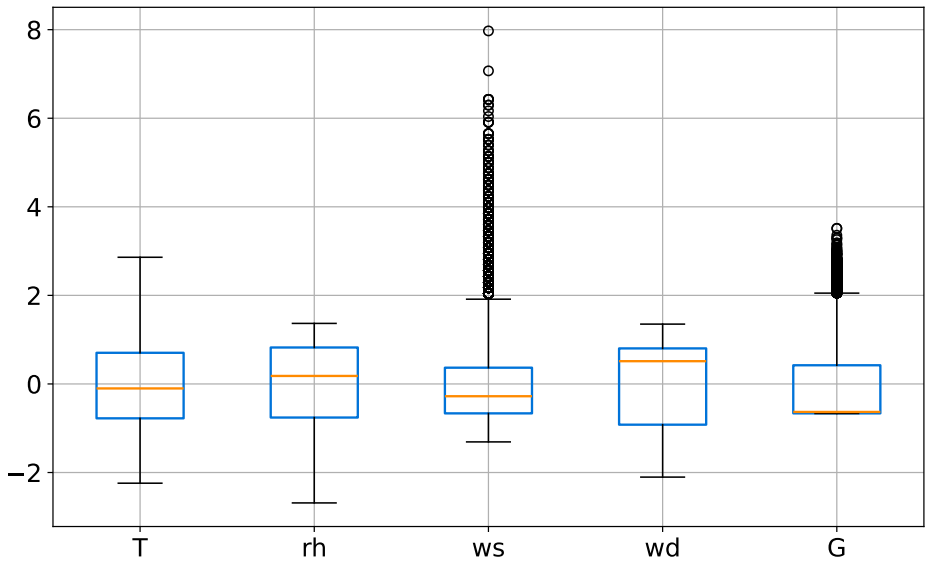
Label	**Features**					
PPV (kW/kWp)	G (W/m^2)	T (°C)	rh (%)	ws (m/s)	wd (°)	timestamp
0.567	400	19.6	57	2.0	250	2022-09-11 15:30
0.670	480	19.5	58	2.4	252	2022-09-11 15:45
0.456	320	19.4	59	2.6	254	2022-09-11 16:00
0.521	380	19.4	61	2.8	256	2022-09-11 16:15

To characterize the variability of the meteorological variables and to illustrate the impact of standardization on the training process, Figs. 2.3a and 2.3b report the distributions of

the raw (non-standardized) and standardized features, respectively. The boxplots highlight the heterogeneity of the measurement ranges (e.g., irradiance up to ∼1200 W/m^2, wind speed typically below 10 m/s). After standardization, all variables have zero mean and unit standard deviation, which facilitates numerical stability during training. In both cases, a significant number of outliers is observed, especially for irradiance and wind speed, likely due to abrupt weather changes or measurement errors. These outliers are preserved in the dataset in order to maintain the natural variability of operating conditions.



(a) Non-standardized features



(b) Standardized features

Figure 2.3 – Distribution of selected meteorological features in the two scenarios. (a) Non-standardized data showing heterogeneous measurement ranges. (b) Standardized features with zero mean and unit variance, which ensure balanced weighting in the training phase. Outliers, visible especially in irradiance and wind speed, are retained to preserve realistic variability.

An illustrative example covering a seven-day time window for one of the three plants is presented in Fig. 2.4. The upper panel shows the normalized PV power, while the lower panels report irradiance, temperature, and humidity for both scenarios (SA in blue, SB in orange). At a preliminary inspection, the irradiance measured by the on-site pyranometer (SA) appears more strongly correlated with PV power than the irradiance retrieved from the satellite API (SB). Although both sources have the same temporal resolution of 15 minutes, the satellite-derived series appear visually smoother, reflecting the spatial averaging and retrieval processes typical of satellite products.

To complement this qualitative assessment with a quantitative metric, the Pearson correlation coefficient r is employed, as it is one of the most widely used statistical tools to evaluate linear dependence between two variables in renewable energy forecasting. The coefficient is defined as:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\,\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}, \tag{2.3}$$

where X and Y represent the two variables under consideration (here, PV power and irradiance), while $\bar{X}$ and $\bar{Y}$ are their sample means. The coefficient r ranges from -1 to $+1$, with values close to $+1$ indicating strong positive correlation, values close to 0 indicating no linear relationship, and values close to -1 indicating strong negative correlation.

Although alternative measures of correlation exist, such as Spearman's rank coefficient or Kendall's tau, which capture monotonic rather than strictly linear relationships, Pearson's r was adopted in this study. The choice is motivated by the fact that the relationship between PV power and irradiance is expected to be predominantly linear, and Pearson's coefficient is the most widely employed metric in the energy forecasting literature to quantify the dependence between meteorological variables and power output.

Applying Eq. (2.3) to normalized PV power and irradiance highlights the differences between the two scenarios. In SA, the correlation reaches $r = 0.92$, reflecting a very strong linear relationship between pyranometer data and PV output. In SB, the correlation decreases to $r = 0.70$, still positive but weaker, consistent with the smoothing effect inherent to satellite-derived data. Nevertheless, both scenarios reproduce the main diurnal patterns of solar resource variability. The relative mean differences between SA and SB amount to 3.6% for irradiance, 0.1% for temperature, and 1.5% for humidity.

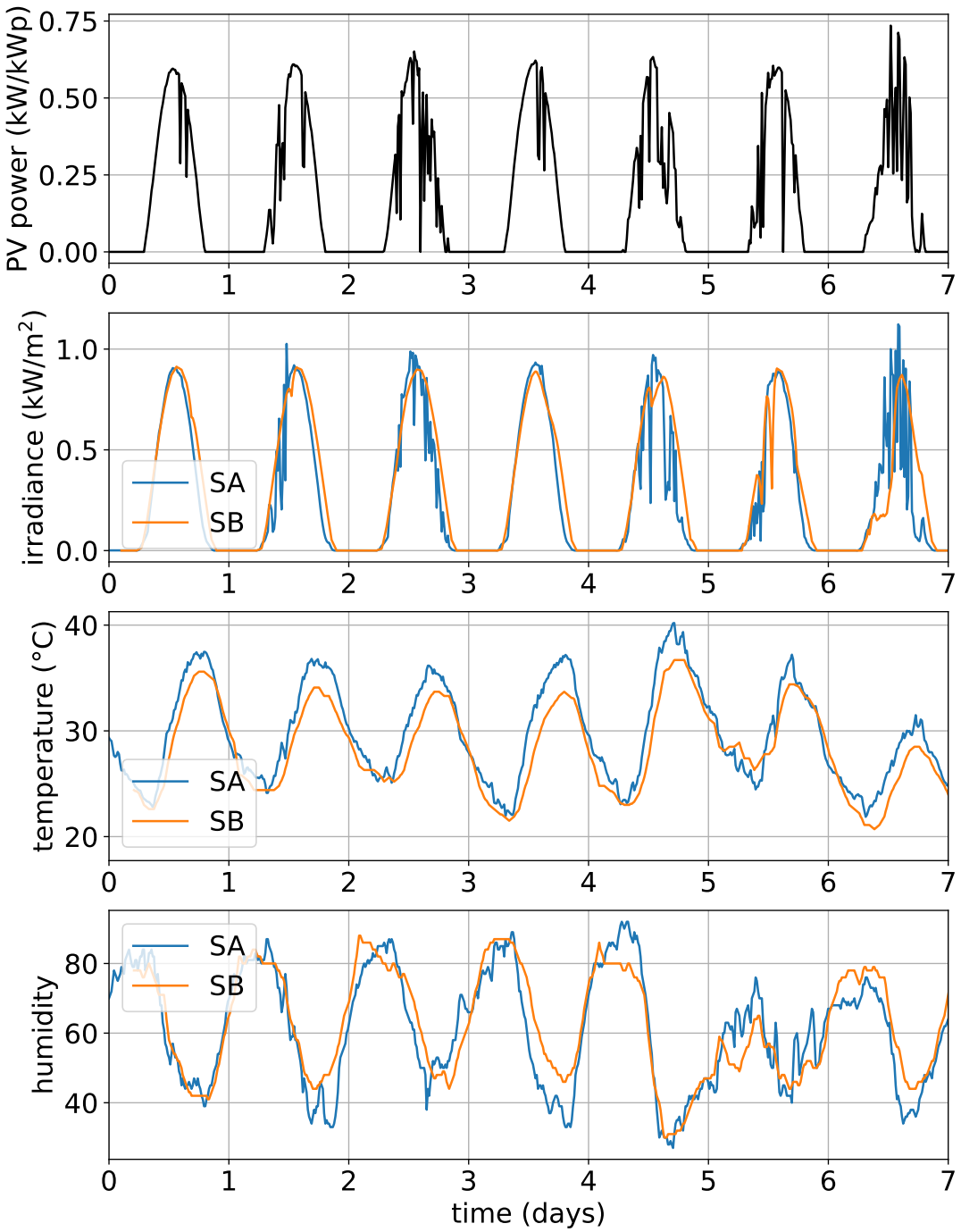


Figure 2.4 – Comparison of normalized PV power (top) and selected meteorological features (irradiance, temperature, and humidity, bottom) for a seven-day window. Blue: local measurements (SA). Orange: satellite-based API data (SB). The figure illustrates the higher correlation between local irradiance and PV power, compared to the smoother satellite-derived series.

Overall, these results emphasize the operational trade-off at the core of this chapter: local measurements (SA) maximize accuracy and correlation with actual PV output but require dedicated instrumentation and are only available to plant operators, while satellite data (SB) offer accessibility and scalability across sites, albeit with a weaker linear relationship to PV production.

2.4 Results

This section presents the outcomes of the forecasting models applied to the three PV plants under investigation. The evaluation was conducted on a test set comprising 72 days, randomly distributed across the year so as to encompass heterogeneous seasonal and meteorological conditions.

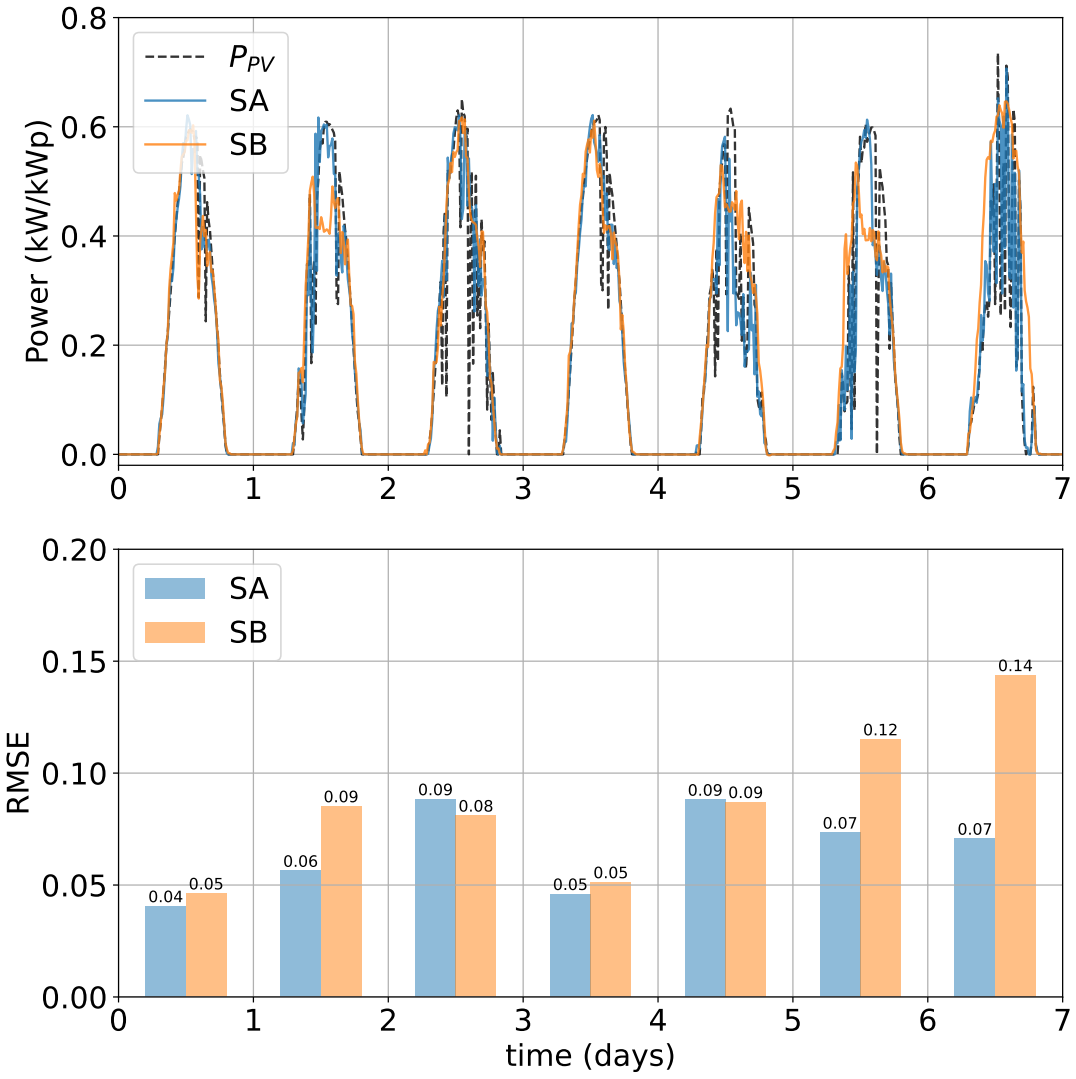


Figure 2.5 – Comparison of measured and predicted power during the representative summer week.

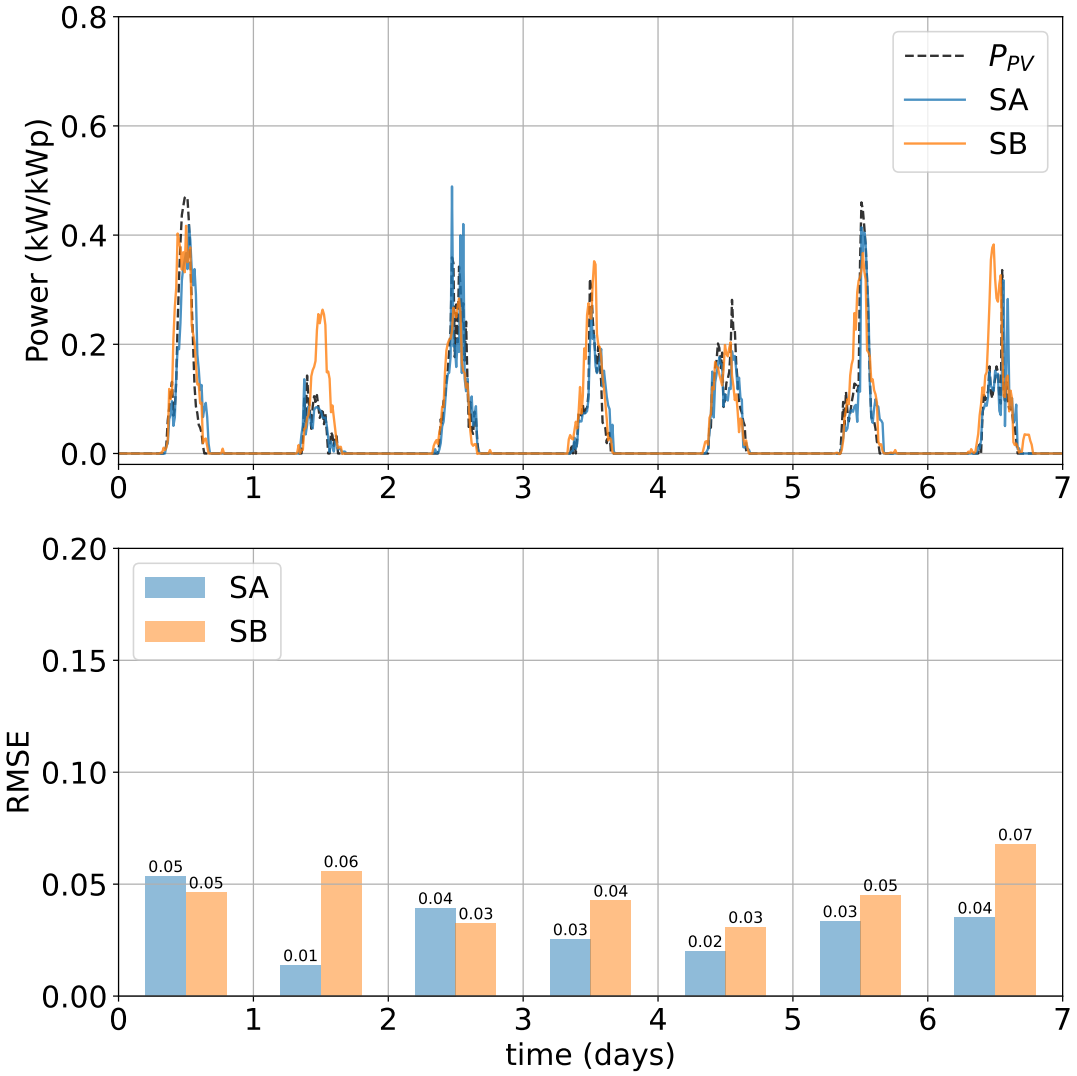


Figure 2.6 – Comparison of measured and predicted power during the representative winter week.

Figures 2.5 and 2.6 illustrate the comparison between measured power and the predictions of Scenario A (SA) and Scenario B (SB) for two representative weeks, one in summer and one in winter. The evidence indicates that SA systematically yields a closer alignment with the measured profiles, particularly on days characterized by pronounced fluctuations in generation. In the summer week, for instance, SB tends to underestimate the midday peaks (notably on days 2, 5, and 6), whereas SA succeeds in reproducing the observed trajectories more faithfully. During the winter week, the second day reveals a marked overestimation by SB, which is effectively corrected in SA, resulting in a curve that better mirrors the measured output.

The daily RMSE values, reported in the lower panels of Figures 2.5 and 2.6, further confirm that forecasting errors scale with the magnitude of the generated energy. In the summer period, when irradiance and production are higher, errors naturally increase, while in winter

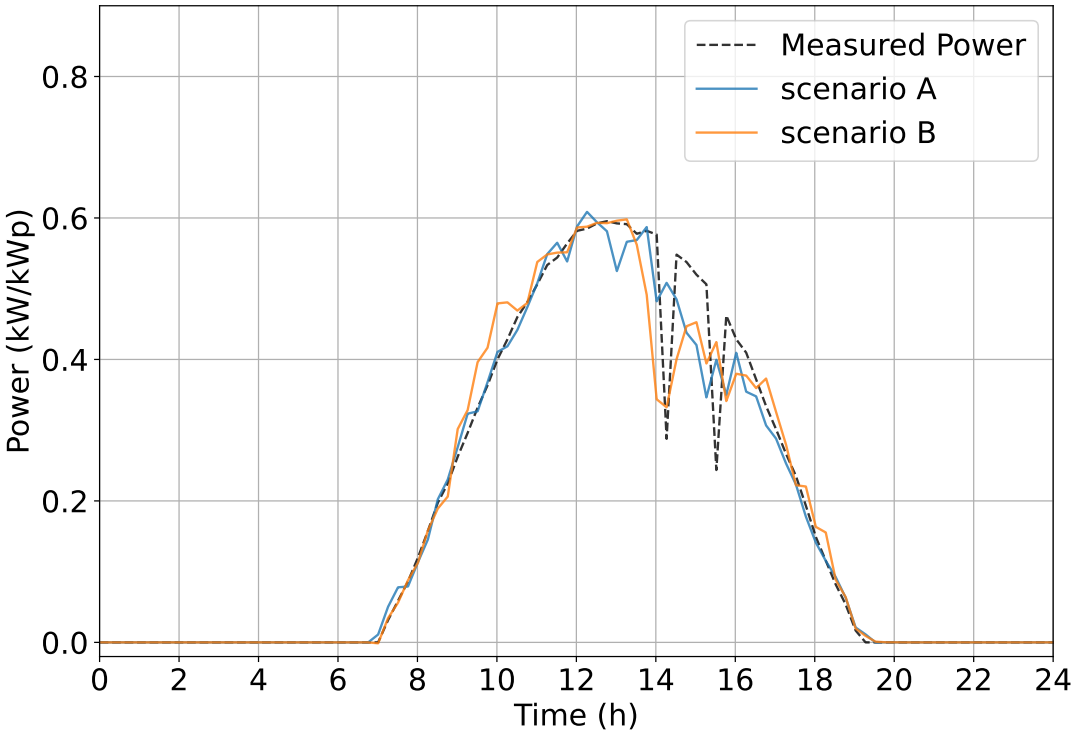


Figure 2.7 – Comparison of measured power and forecasts from the two scenarios for a representative test day.

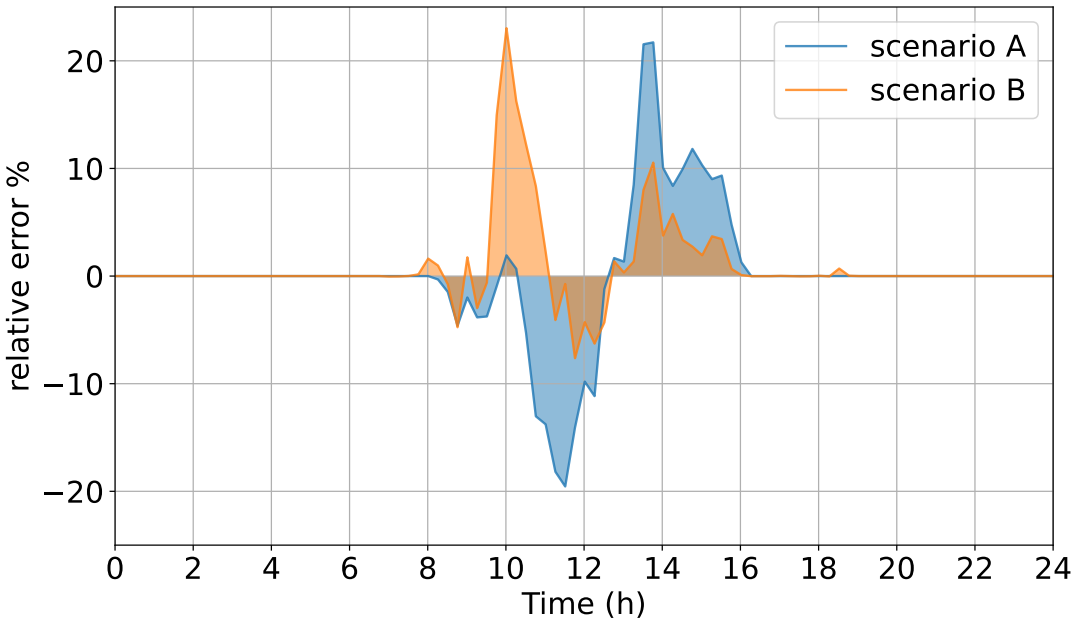


Figure 2.8 – Relative error between measured and predicted power for the same representative day.

Table 2.2 – Overall RMSE and R^2 values for each PV plant under both forecasting scenarios.

PV plant	Scenario A		Scenario B	
	RMSE	R^2	RMSE	R^2
PV plant 1	0.049	0.945	0.056	0.927
PV plant 2	0.038	0.968	0.048	0.941
PV plant 3	0.049	0.946	0.064	0.897

they remain significantly smaller. Across both weeks, SA consistently attains lower RMSE values than SB, with the difference being most pronounced under unstable meteorological conditions, where rapid cloud transients induce abrupt fluctuations in production. By contrast, in more stable conditions the gap narrows, and in certain instances SB even achieves comparable or slightly superior performance.

A more granular assessment is provided in Figures 2.7 and 2.8, which refer to a representative single day of the test set. As shown in Figure 2.7, SA follows the measured profile with remarkable fidelity throughout the entire day, while SB exhibits larger discrepancies, particularly during midday when production reaches its maximum. Both models display a slight underestimation during the early morning and late afternoon, although the deviation remains marginal.

The corresponding relative error is depicted in Figure 2.8. Scenario A maintains consistently smaller deviations over the full time horizon, whereas Scenario B reveals more pronounced fluctuations, especially during the central hours of the day. These results reinforce the findings of the weekly analyses and highlight the ability of SA to capture intraday variability with greater accuracy.

Finally, Table 2.2 summarizes the aggregated performance across all 72 days and the three PV plants. Scenario A achieves maximum RMSE values of 0.049 and R^2 values as high as 0.968, while Scenario B records a maximum RMSE of 0.064 and R^2 up to 0.941. Although the superiority of SA is evident, the absolute difference in RMSE between the two approaches remains relatively limited (below 0.02), thus indicating that SB can still provide reasonably accurate forecasts despite relying solely on satellite-based data. Moreover, SB offers an important practical advantage: meteorological information can be retrieved straightforwardly via API, without requiring on-site weather stations, thereby facilitating large-scale deployment. However, SA clearly demonstrates superior accuracy and robustness in capturing both seasonal and intraday variability.

2.5 Chapter summary

This chapter has investigated the influence of meteorological data sources on the performance of photovoltaic power forecasting models, thereby addressing a significant gap in the literature. While numerous studies have demonstrated the effectiveness of machine learning techniques such as gradient boosting for renewable energy forecasting, comparatively little attention has been paid to the systematic evaluation of how different sources of meteorological information, namely local, in-situ measurements versus satellite-based data, shape forecasting accuracy. This lack of comparative evidence is particularly relevant from an operational perspective, as it limits the ability of both researchers and practitioners to select forecasting approaches that are not only accurate but also practical and scalable.

The work presented here has sought to bridge this gap through a structured comparison of two forecasting scenarios. In Scenario A (SA), the gradient boosting model was trained using high-resolution data collected directly from local weather stations and pyranometers installed near the PV plants. In Scenario B (SB), the same set of meteorological features was retrieved from satellite-based services via API, thus ensuring broad accessibility but at the expense of lower spatial and temporal resolution. Both scenarios were evaluated using one year of data from three photovoltaic plants of different sizes and geographical locations across Italy, thereby ensuring that the assessment encompassed heterogeneous environmental and operational conditions.

The comparative analysis revealed clear differences. Scenario A consistently outperformed Scenario B, achieving RMSE values as low as 0.038 (with maxima of 0.049) and R^2 values up to 0.968, compared to a maximum RMSE of 0.064 and R^2 values up to 0.941 for Scenario B. These results confirm that access to site-specific meteorological measurements provides a measurable advantage in predictive accuracy. At the same time, Scenario B demonstrated that even in the absence of local instrumentation, reliable forecasts can still be generated, with errors remaining within an acceptable range for operational use.

Beyond the numerical results, the study highlights an important trade-off between accuracy and accessibility. Scenario A offers superior precision but relies on instrumentation that is costly to install and maintain, and often unavailable outside the domain of plant operators. Scenario B, in contrast, leverages universally accessible satellite-based data that can be retrieved automatically via API, making it a scalable solution for multi-site or large-area forecasting applications, even though it entails a modest reduction in accuracy.

By systematically quantifying this trade-off under a unified methodological framework, the chapter makes an original contribution to the PV forecasting literature. It demonstrates that the choice of meteorological data source is as decisive as the choice of modeling technique itself, and that careful consideration of this factor is essential for the development of forecasting tools that are both technically sound and operationally deployable.

Chapter 3

PV power forecasting based on real weather forecast data and Synthetic Feature Engineering

3.1 Motivation and objectives

The accuracy of photovoltaic power forecasting models is strongly influenced by the quality, reliability, and availability of the meteorological inputs. In operational contexts, input data are rarely derived from on-site measurements such as pyranometers or dedicated weather stations, since these require investment, maintenance, and continuous calibration. More commonly, the data are obtained from heterogeneous sources, including terrestrial weather stations distributed across the territory or satellite-based datasets made accessible through online platforms and APIs. Consequently, the predictive performance of a model is not only determined by the adopted forecasting algorithm, but also by the resolution, coverage, and inherent uncertainty of the meteorological information on which it relies. This introduces a fundamental trade-off between accuracy and accessibility that plays a central role in the design of forecasting methodologies [117].

Among the available meteorological variables, solar radiation represents the dominant factor affecting PV production. Forecasts of solar radiation are typically obtained by combining clear-sky GHI with cloud cover information derived from satellite observations. However,

the algorithms employed to merge these components are often proprietary or not fully documented, which limits transparency and hinders reproducibility. Moreover, the intrinsic uncertainty of cloud dynamics translates into significant errors in solar radiation forecasts, and consequently in PV power predictions. The lack of direct access to accurate and real-time irradiance data thus constitutes one of the main bottlenecks in the development of reliable forecasting tools.

In recent years, most machine learning models for PV forecasting have been trained on historical power and meteorological datasets [98]. While this approach has enabled good performance during training and retrospective validation, it suffers from a fundamental limitation: in operational practice, models do not receive historical data as inputs, but rather weather forecasts. Consequently, the exclusive reliance on historical records may generate a mismatch between training conditions and real-world operational inputs, ultimately reducing the reliability of predictions when models are applied to actual forecast data[101–103]. This mismatch between training data and operational inputs has motivated a growing number of contributions that explicitly incorporate forecasted meteorological data into PV prediction models. For example, [118] introduced the SolarPredictor architecture, designed to account for the uncertainty of weather forecasts and to enhance accuracy in real-world conditions, while [119] proposed a hybrid approach combining Long Short-Term Memory (LSTM) networks with self-attention mechanisms, achieving a 15.8% improvement over conventional LSTM baselines. These works illustrate the importance of shifting the focus from historical records to actual forecast data in order to strengthen the operational relevance of PV forecasting models.

To overcome these challenges, different strategies have been proposed in the literature. A common approach is to apply feature engineering techniques, where new variables are created from the available data in order to capture latent patterns or to compensate for missing information. Previous studies have focused on statistical transformations such as moving averages, lagged values, or dimensionality reduction techniques like PCA. For instance, in [68], an LSTM-based model integrated synthetic weather forecasts with historical irradiance and sky-type classifications, achieving a 44.6% improvement in accuracy over daily forecasts. Nevertheless, the reliance on historical patterns may reduce adaptability to unexpected weather changes. Similarly, in [69], a mathematical model identified twelve key meteorological factors and analyzed them through PCA to improve solar energy forecasts. While effective in highlighting correlations, this approach is constrained by the dependence

on predefined models, which may limit flexibility in capturing the complex dynamics of PV power generation.

Other works have explored alternative sources of meteorological information and advanced feature engineering methods. In [54], sky cover indices derived from total sky imagers, infrared radiation, and GHI were used as inputs to an artificial neural network, showing that integrating diverse data sources can significantly enhance solar radiation forecasts. However, such methods require specialized instruments and infrastructures that are not always available for large-scale or distributed applications. In [70], feature engineering techniques were applied by introducing lagged power values, moving averages, and time-based variables (e.g., hour of the day, day of the year), which were then used in a stacked ensemble learning model combining Random Forest, XGBoost, and Multiple Linear Regression. This demonstrated the potential of combining statistical transformations with ensemble methods to increase forecasting accuracy. More recently, CNNs have been employed for automatic feature extraction, as in [71], where they were successfully applied to identify faults in PV panels using image data, further highlighting the versatility of data-driven approaches.

Despite these advances, most existing studies address synthetic features and feature engineering with the goal of refining available meteorological variables or incorporating historical PV data. Few works explicitly tackle the scenario in which some key meteorological variables, such as real-time solar irradiance, are entirely unavailable. This gap is particularly relevant in practice, since real-time measurements are rarely accessible to operators who rely on public APIs. In such cases, the question arises whether synthetic features, derived by combining more accessible variables (e.g., clear-sky GHI and cloudiness indices), can act as substitutes for unavailable measurements while still preserving forecast accuracy.

The present chapter addresses these open questions by investigating photovoltaic power forecasting based on real weather forecast data and by evaluating the role of synthetic features in scenarios of limited data availability. Two distinct forecasting models are designed and compared in this study:

- **Base-mode:** This reference model is trained using the set of commonly available meteorological forecast features, including solar radiation, clear-sky GHI, cloud cover, precipitation, relative humidity, temperature, UV index, and weather rank. It does not employ any synthetic features and thus reflects the typical operational setup of

PV forecasting when direct irradiance-related inputs are accessible. The Base-model serves as a benchmark for evaluating the added value of synthetic features.

- **SF-model:** In this model, solar radiation, clear-sky GHI, and cloud cover are removed from the feature set and replaced by a single synthetic feature at a time (from χ_1 to χ_5). Each synthetic feature is constructed to approximate the nonlinear interactions between atmospheric variables and irradiance dynamics. The SF-model therefore represents the scenario in which some meteorological variables are unavailable, testing whether synthetic features can act as suitable substitutes while maintaining forecasting accuracy and reducing feature redundancy.

The model performance is evaluated using the RMSE and the R^2.

The main goal of this analysis is therefore twofold: first, to assess the potential of synthetic features as substitutes for unavailable meteorological variables; and second, to evaluate the predictive capabilities of models trained and tested directly on real forecast data, rather than historical records. By explicitly considering both aspects, the chapter aims to provide a comprehensive perspective on the robustness of machine learning approaches for PV forecasting under realistic, operationally relevant conditions.

3.2 Data analysis and model setup

This section presents both the description of the datasets employed in this study and the methodological setup of the forecasting models. The analysis is structured in three parts: first, the photovoltaic production data and corresponding meteorological forecasts are introduced, together with the additional parameter age, which accounts for forecast reliability. Second, the set of weather-related predictors is analyzed, with emphasis on the selection of the most relevant features and on the design of synthetic variables aimed at compensating for the lack of direct irradiance measurements. Finally, the machine learning framework adopted for forecasting, based on the LGBM, is described in detail, including training procedures, hyperparameter tuning, and evaluation metrics. This integrated approach provides a unified perspective of the entire forecasting workflow, ensuring consistency between input information and predictive methodology.

3.2.1 Dataset description

This section provides a detailed description of the datasets employed for model training, with specific attention to both photovoltaic power generation and forecasted meteorological variables.

With regard to photovoltaic production, the data were collected from a plant located in Italy and cover the period from November 2022 to June 2023, with a temporal resolution of 15 minutes. Such granularity enables the capture of intraday dynamics in generation, which are often characterized by rapid fluctuations caused by transient phenomena such as cloud passages.

The meteorological dataset includes the corresponding weather forecasts for the same time span, also provided with a 15-minute resolution. Forecast data were retrieved from an online platform [116], which offers access to atmospheric variables generated on the basis of the geographic coordinates of the plant site. This approach reproduces the actual operational context in which forecast data are typically obtained by plant operators or grid managers.

To ensure comparability across series, the PV power values were normalized with respect to the rated peak capacity of the plant (kW/kWp). The comprehensive dataset was constructed by downloading, for each day, the meteorological forecasts for the subsequent 24 hours, maintaining the same 15-minute time step.

A key element introduced in this work is the parameter referred to as age, which indicates the time elapsed since the forecast was issued. This parameter reflects a well-established concept in weather forecasting: more recent forecasts (younger age) are generally more accurate, as they are based on up-to-date input data, whereas older forecasts (older age) tend to be less reliable due to the inherently dynamic and evolving nature of atmospheric conditions. Including this information enriches the dataset with a temporal dimension of forecast reliability, thereby enhancing the realism of the operational scenario.

In summary, the structuring of the dataset and the introduction of the age parameter provide a coherent and chronologically consistent framework of information, ensuring a solid foundation for the training and validation of photovoltaic power forecasting models.

3.2.2 Standard Weather Features

The selection of appropriate meteorological variables plays a central role in the development of robust PV forecasting models. While machine learning methods can, in principle, process a large number of inputs, the inclusion of weakly relevant variables may introduce redundancy and noise, leading to overfitting and a reduced ability to generalize to unseen data [120]. For this reason, an initial feature selection step was carried out using Pearson's correlation coefficient, calculated between each available meteorological variable and the measured PV power output.

The heatmap in Fig. 3.1 provides a preliminary quantitative assessment of the strength and direction of linear dependencies: positive correlations close to +1 (in red) indicate variables that tend to increase with PV generation, whereas negative correlations close to –1 (in blue) highlight inverse relationships.

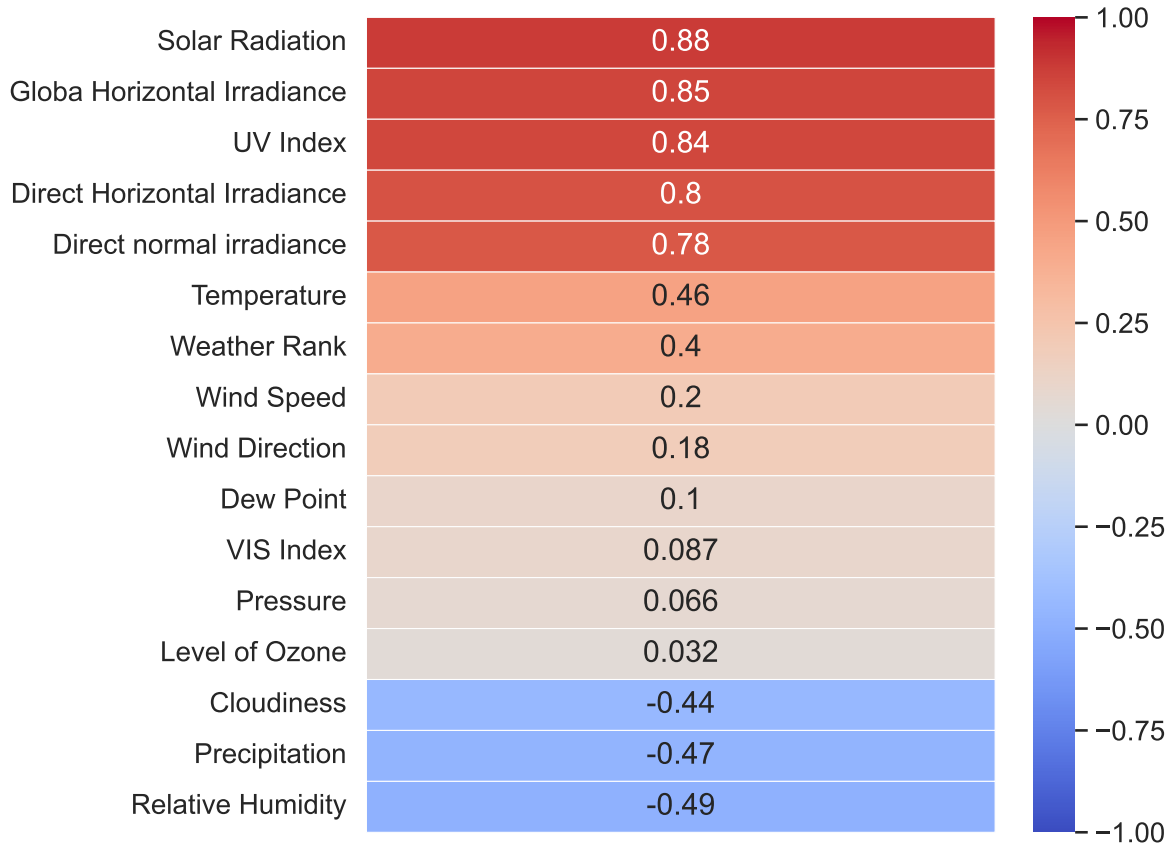


Figure 3.1 – Heatmap of Pearson's correlation coefficient between meteorological variables and PV power production. Image retrieved from [121].

In addition to meteorological variables with a direct and well-known physical link to solar generation, such as solar radiation or GHI, the dataset also includes derived indicators that capture broader atmospheric conditions. These include the *weather rank*, a synthetic score ranging from 0 to 100 that summarizes the overall meteorological situation; precipitation, expressed on a 0–8 scale, reflecting rainfall intensity; the UV index (0–11), representing the strength of ultraviolet radiation at the surface; and the VIS index, a dimensionless measure of atmospheric visibility. Cloudiness, a particularly relevant factor for solar generation, is expressed as the percentage of sky covered by clouds (0 = clear sky, 100 = overcast).

Based on the correlation analysis, only variables with an absolute correlation greater than 0.4 were retained, resulting in the following set: air temperature (°C), precipitation (mm), relative humidity (%), weather rank, UV index, solar radiation (W/m^2), global horizontal irradiance (GHI, W/m^2), and cloudiness. Among these, solar radiation, the UV index, and GHI exhibit the strongest positive correlation with PV output, as expected from the underlying physics of photovoltaic conversion. Conversely, relative humidity and cloudiness show marked negative correlations, since high humidity and dense cloud cover typically attenuate the solar irradiance reaching the panels.

It is important to note that correlation values should not be interpreted solely as causal relationships, since meteorological variables are often interdependent. For example, solar radiation and GHI are strongly collinear, while cloudiness is inherently linked to humidity and precipitation. Nevertheless, correlation analysis provides a practical criterion for reducing dimensionality while retaining the most informative variables for PV forecasting.

3.2.3 Syntetic features

A central contribution of this work is the introduction of synthetic features, which are designed to act as substitutes for unavailable irradiance data and to enrich the set of predictors available for photovoltaic forecasting. The motivation for their development arises from a fundamental practical limitation: while irradiance is the single most influential variable for PV power generation, direct ground-based measurements are rarely available in operational settings. Instruments such as pyranometers or all-sky imagers are expensive to install, require regular calibration, and are typically deployed only in research facilities or large demonstration plants. In contrast, other atmospheric variables such as cloud fraction and clear-sky global GHI can be accessed at no cost through public APIs and satellite-based

datasets. These variables, although indirect, capture essential aspects of solar resource availability and are therefore suitable candidates for constructing alternative predictors.

The key idea behind synthetic features is to combine cloud cover information with theoretical clear-sky irradiance in order to approximate the effect of atmospheric conditions on solar radiation at ground level. By embedding non-linear functional forms, these features aim not only to mimic physical attenuation but also to reflect the abrupt and highly variable impact that clouds can have on solar radiation. In fact, the relationship between cloudiness and irradiance is far from linear: minor fluctuations in cloud fraction may cause sharp irradiance reductions due to scattering, absorption, and cloud-edge effects, while in other situations even significant increases in cloud cover may lead to relatively modest irradiance changes. Capturing this non-linearity is thus crucial for improving forecast accuracy.

Within this framework, five synthetic features (χ_1–χ_5) were developed by combining clear-sky GHI (G) and cloud cover fraction (Cl). The first four (χ_1–χ_4) each represent a different functional form of attenuation, designed to approximate various possible dynamics of cloud–radiation interaction [122, 123]. The fifth feature (χ_5) provides a more flexible hybrid formulation, explicitly combining both irradiance and cloudiness into a single weighted index.

The first synthetic feature is defined as:

$$\chi_1 = G \cdot \left(\frac{100 - Cl}{100} \right)^{\beta_1}. \tag{3.1}$$

This expression implements an exponential-type attenuation, where the irradiance is reduced proportionally to the fraction of clear sky. The coefficient β_1 controls the degree of sensitivity, allowing the reduction to be either gentle or steep. The rationale is consistent with atmospheric transmittance models, which predict that the optical path of solar radiation decreases as cloud coverage increases [124].

The second feature introduces a subtractive penalty:

$$\chi_2 = G - \left(\frac{G \cdot Cl}{100} \right)^{\beta_2}. \tag{3.2}$$

Here, irradiance is expressed as the difference between its theoretical clear-sky value and a penalty term dependent on both cloudiness and GHI. Unlike χ_1, which applies a proportional scaling, χ_2 allows for stronger non-linear effects in which small variations in

cloud cover lead to disproportionately large reductions in irradiance. This is consistent with empirical observations of intermittent and rapidly varying cloud patterns, which can cause sharp fluctuations in PV output [125].

The third synthetic feature applies an exponential decay:

$$\chi_3 = G \cdot e^{-(\alpha \cdot Cl/100)}. \tag{3.3}$$

The exponential form reflects the fact that irradiance attenuation tends to increase progressively and non-linearly with cloud cover. The parameter α defines the steepness of this decay. This feature is particularly suited to reproduce scenarios in which irradiance decreases moderately under partial cloudiness but drops very rapidly once a critical fraction of sky is obscured.

The fourth feature assumes a quadratic relationship:

$$\chi_4 = G - \gamma \left(\frac{Cl}{100}\right)^2. \tag{3.4}$$

The quadratic term accentuates the impact of high cloud fractions. At low levels of cloudiness, irradiance is only marginally reduced, but as coverage approaches full sky obstruction, the reduction becomes much more severe. This form is motivated by studies reporting that the effect of clouds often accelerates under dense or layered cloud formations [122].

The last synthetic feature, χ_5, is conceptually different from the previous ones because it is not strictly an attenuation model but a weighted combination of two complementary indicators: normalized clear-sky irradiance and the fraction of clear sky:

$$\chi_5 = w_1 \cdot \frac{G}{G_{\max}} + (1 - w_1) \cdot \frac{100 - Cl}{100}. \tag{3.5}$$

Here, the parameter w_1 determines the balance between the contribution of irradiance and cloud fraction. This formulation acknowledges that the importance of irradiance and cloudiness may vary depending on meteorological conditions: in some situations irradiance is the dominant driver, while in others the dynamics of cloud cover are more critical. By adjusting w_1, the feature adapts to these different contexts, offering greater flexibility.

The parameters β_1, β_2, α, γ, and w_1 were calibrated using a grid search over the interval [0, 1] with a step size of 0.1. For each candidate value, the forecasting model was retrained

and evaluated on the validation dataset. The RMSE was chosen as the performance metric, and the parameter producing the lowest RMSE was selected as optimal. This systematic calibration ensured that the synthetic features were not only theoretically plausible but also empirically effective in reducing forecast errors. Importantly, the grid search allowed the analysis of how small variations in parameter values influence model performance, thus providing insight into the robustness of the proposed features.

To complement this analysis, Pearson's correlation coefficient was computed between each synthetic feature and the measured PV power. The results, reported in Table 3.1, reveal that all five features exhibit strong positive correlations with PV output, with χ_3 achieving the highest value (0.886), followed closely by χ_5 (0.860) and χ_4 (0.859).

Table 3.1 – Pearson's correlation coefficients between synthetic features and measured PV power.

Synthetic feature	Pearson's Correlation Coefficient
χ_1	0.844
χ_2	0.843
χ_3	0.886
χ_4	0.859
χ_5	0.860

Although correlation does not directly measure predictive accuracy, this analysis provides a useful first indication of the explanatory capacity of the synthetic features. The superior performance of χ_3 suggests that exponential attenuation functions may capture irradiance dynamics more effectively than simple scaling or quadratic models, while the competitive results of χ_5 highlight the potential of hybrid formulations that balance different atmospheric drivers. Overall, these findings confirm the feasibility of constructing effective substitutes for unavailable irradiance measurements by leveraging accessible data sources, thereby enhancing the realism and robustness of PV forecasting models.

Although correlation does not directly measure predictive accuracy, this analysis provides a useful first indication of the explanatory capacity of the synthetic features. The superior performance of χ_3 suggests that exponential attenuation functions may capture irradiance dynamics more effectively than simple scaling or quadratic models, while the competitive results of χ_5 highlight the potential of hybrid formulations that balance different atmospheric drivers. Overall, these findings confirm the feasibility of constructing effective substitutes for unavailable irradiance measurements by leveraging accessible data sources, thereby enhancing the realism and robustness of PV forecasting models.

3.2.4 Model setup

The forecasting algorithm selected for this study is the LGBM, chosen for its training efficiency, scalability, and strong generalization capabilities. Compared to other ensemble methods such as XGBoost and CatBoost, LGBM offers faster convergence and more efficient handling of large datasets [126]. These properties are particularly valuable in PV forecasting applications, where data have high temporal resolution and must be updated frequently.

A distinctive feature of LGBM is its leaf-wise tree growth strategy, in contrast to the traditional level-wise approach. At each iteration, the algorithm grows the leaf that results in the greatest reduction of the loss function. This produces deeper and more accurate trees with fewer iterations, improving both training speed and predictive accuracy [127]. Although this approach is more aggressive, it is balanced by regularization parameters such as maximum number of leaves, maximum depth, and minimum child weight, which help to control overfitting and ensure good generalization.

In addition, LGBM implements histogram-based decision tree learning, which discretizes continuous variables into bins. This significantly reduces both computation time and memory usage, making LGBM particularly suitable for large-scale datasets and near real-time applications such as PV power forecasting for grid management. Another advantage is its ability to handle both numerical and categorical variables efficiently, reducing the need for complex preprocessing. Furthermore, its internal feature selection mechanisms automatically reduce the importance of less relevant variables, which improves model performance while shortening training times.

In this study, the target variable (label) is the measured PV power output, while the input features correspond to the meteorological variables most correlated with PV production. The dataset was divided into three subsets: 80% was used for training, randomly sampled to allow the model to learn the relationships between inputs and outputs; 20% was used for validation, also randomly sampled but distinct from the training set; and a separate test set was created from seven consecutive days, entirely excluded from the previous stages, in order to evaluate the generalization capability of the model on independent and temporally non-overlapping data.

Hyperparameter tuning was performed using a random search strategy, implemented with the RandomizedSearchCV function [128, 129]. The search space included: boosting methods (GBDT and DART), number of leaves between 31 and 50, learning rate between 0.01 and

0.19, feature and bagging fractions between 0.5 and 0.9, bagging frequency between 1 and 6, L1 and L2 regularization values between 0.0 and 1.0, minimum child weight between 0.001 and 0.999, and maximum depth between 5 and 12.

To mitigate overfitting, early stopping was applied with a tolerance of 50 rounds, interrupting training when no improvement in validation RMSE was observed. In addition, 5-fold cross-validation was employed to provide a more robust evaluation and to improve the model's generalization capability.

The metrics used to assess model performance are the root mean square error (RMSE), expressed in W/kWp, and the coefficient of determination (R^2). The entire workflow, including data preprocessing, model training, and validation, was implemented in Python using the Scikit-learn library [128]. The loss function used during hyperparameter tuning was the negative mean squared error, which is particularly suitable for regression tasks since it penalizes larger deviations between predicted and observed values.

3.3 Results

The models developed in this study were evaluated over the test week from May 8 to May 15, 2023. This period was selected because it includes days with stable and sunny weather, where the production curve is regular, as well as days characterized by strong variability due to cloud passages. Such heterogeneity provides a representative benchmark to assess both the accuracy of the models under favorable conditions and their robustness when weather conditions change abruptly.

Figure 3.2 illustrates the comparison between the measured PV power and the predictions obtained with the Base-model. Overall, the predictions follow the real production trend closely. On variable days, such as May 9 and May 10, the model is able to capture the main fluctuations, although discrepancies appear at some peaks, likely due to rapid cloud dynamics not fully represented by the forecasted inputs. On more stable days, such as May 12 and May 14, the predictions are much closer to the measured values, with only a slight tendency to underestimate peak power. The performance metrics confirm these observations, with an RMSE of 87.232 and an R^2 of 0.875.

A preliminary analysis conducted on the validation dataset allowed the identification of the optimal weights for the five synthetic features, i.e., β_1, β_2, α, γ and $w1$ in order to minimize

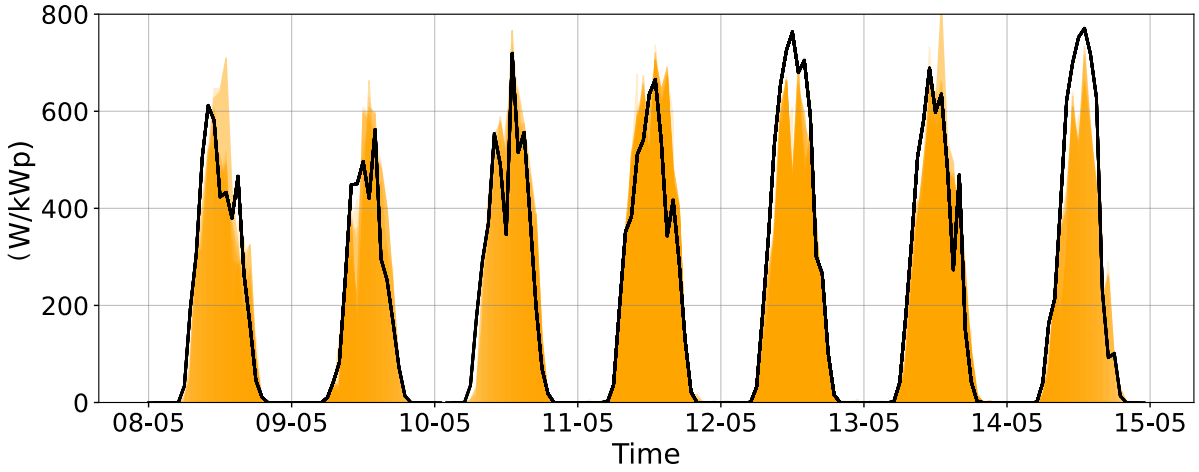


Figure 3.2 – Comparison between measured power (black line) and predicted power (orange area) by the Base-model during the reference week (May 8–15, 2023). Image retrieved from [121]

the RMSE and maximize the R^2 value. The minimum RMSE and corresponding R^2 values versus the weights for the analyzed synthetic features is reported in Table 3.2.

The best RMSE values are associated with the synthetic features χ_2 and χ_5. The lowest RMSE, 19.05, was achieved with χ_5 using a weight $w_1 = 0.9$. Conversely, the feature with the worst performance is χ_1, which resulted in an RMSE of 24.337 and an R^2 value of 0.985.

This preliminary analysis highlights the ability of the model trained on synthetic features to accurately predict the measured power. Furthermore, it allows the identification of the most suitable feature for optimizing PV power forecasting. This analysis indicated that χ_5 and χ_1 exhibit, respectively, the most accurate and least accurate results, in terms of RMSE and R^2. Therefore, the models based on these features were analyzed for the reference week.

Figure 3.3 shows the results of the SF-model trained with the synthetic feature χ_5, identified in the preliminary analysis as the most accurate. In this case, the forecasts reproduce the measured curves more faithfully than the Base-model, especially during days with changing

Table 3.2 – Optimal weights for each synthetic feature with the corresponding RMSE and R^2 values.

SF	SF weight	Optimal value	RMSE (W/kWp)	R^2
χ_1	β_1	0.1	24.337	0.985
χ_2	β_2	0.1	19.674	0.990
χ_3	α	0.1	22.686	0.987
χ_4	γ	0.3	19.777	0.990
χ_5	$w1$	0.9	19.050	0.991

weather. For instance, on May 8 and May 13, the model captures variations in production with greater precision, while on May 12 and May 15 (clear-sky days), the agreement with the measured values is excellent, with negligible errors. These improvements are reflected in the performance metrics: the RMSE decreases to 84.013, while the R^2 rises to 0.888, highlighting the superior accuracy and explanatory power of the χ_5-based model.

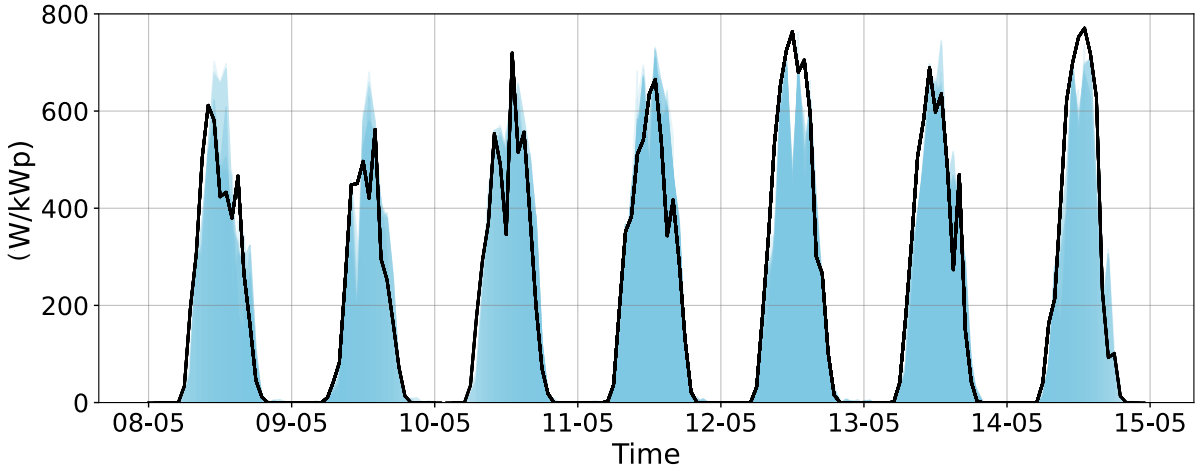


Figure 3.3 – Comparison between measured power (black line) and predicted power (blue area) by the SF-model trained with χ_5 during the reference week. Image retrieved from [121].

Figure 3.4 reports the results for the SF-model trained with χ_1, the least effective synthetic feature. Although the model preserves a reasonable predictive ability, it tends to systematically underestimate PV production during stable days such as May 12 and May 14. On variable days like May 9 and May 11, the model captures only part of the observed fluctuations, missing some peaks. The performance metrics confirm this weaker behavior: RMSE increases to 88.523, while R^2 slightly decreases to 0.874, performing even worse than the Base-model.

Table 3.3 summarizes the performance metrics for the Base-model and all SF-models. The comparison reveals that χ_5 provides the best results, slightly improving upon the Base-model, whereas χ_1 yields the poorest performance. The other synthetic features (χ_2, χ_3, and χ_4) achieve intermediate values, close to those of the reference model.

These results demonstrate that synthetic features can serve as effective substitutes for unavailable irradiance data, yielding forecasts comparable to those obtained with conventional meteorological variables. In particular, the superior performance of χ_5 indicates that hybrid formulations combining irradiance and cloud cover provide a more flexible and realistic representation of atmospheric processes, thereby enhancing the robustness of PV power forecasting under operational conditions.

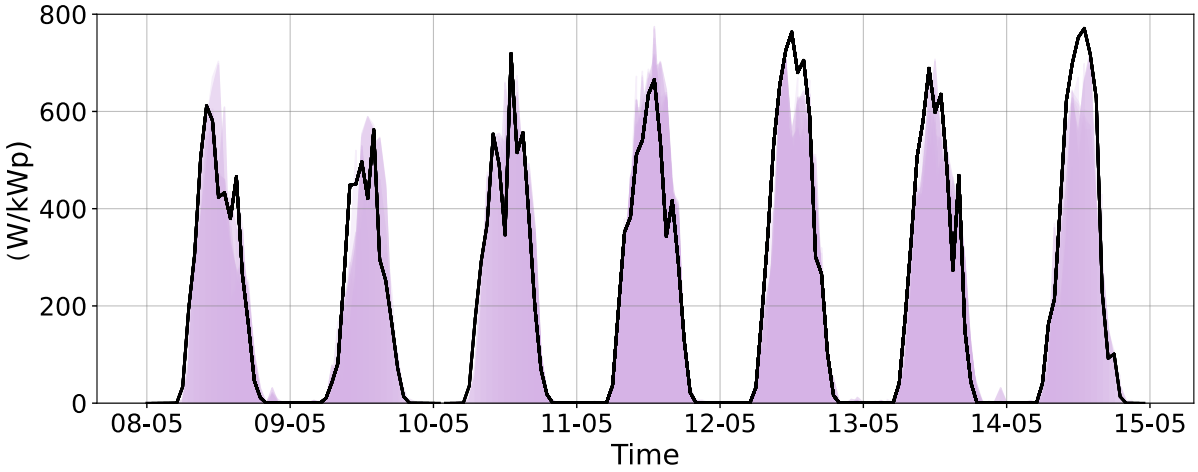


Figure 3.4 – Comparison between measured power (black line) and predicted power (purple area) by the SF-model trained with χ_1 during the reference week. Image retrieved from [121]

Table 3.3 – Comparison of RMSE and R^2 values for the Base-model and SF-models.

Model	**RMSE (W/kWp)**	**R^2**
Base-model	87.232	0.875
SF-model (χ_1)	88.523	0.874
SF-model (χ_2)	85.354	0.883
SF-model (χ_3)	86.169	0.881
SF-model (χ_4)	88.209	0.875
SF-model (χ_5)	84.013	0.888

3.4 Chapter summary

This chapter has examined the extent to which the selection and construction of meteorological inputs influence the performance of photovoltaic power forecasting models under operational conditions. While previous research has highlighted the effectiveness of modern machine learning techniques, particularly gradient boosting algorithms, considerably less attention has been devoted to models trained directly on forecasted rather than historical meteorological data, and to the role of synthetic features in the absence of direct irradiance measurements. This issue is especially relevant in practice, since plant operators and grid managers typically rely on publicly available forecast data provided via APIs rather than on-site pyranometers, and must therefore reconcile accuracy, accessibility, and scalability.

To address this problem, the chapter implemented a unified methodological framework that combines forecast-based training, the explicit handling of the variable age, which quantifies the temporal distance from forecast issuance, and the application of LGB models.

Two forecasting scenarios were constructed and analyzed. In the first, the Base-model, training was performed using conventional meteorological forecast variables, including solar radiation, clear-sky global horizontal irradiance, and cloud cover, together with temperature, humidity, precipitation, UV index, and a composite weather rank. In the second scenario, the SF-model, irradiance-related inputs were removed and replaced one at a time by synthetic features (χ_1–χ_5), designed by combining theoretical irradiance and cloud cover through nonlinear functional forms intended to approximate cloud–radiation interactions.

The evaluation on the test week of 8–15 May 2023 demonstrated that forecast-based training ensures robust accuracy even in the absence of local measurements. The Base-model achieved an RMSE of 87.232 W/kWp and an R^2 of 0.875. When irradiance-related variables were substituted with synthetic features, performance remained comparable and in some cases improved. The SF-model based on the hybrid feature χ_5 attained the best performance, with an RMSE of 84.013 W/kWp and $R^2 = 0.888$, whereas χ_1 produced a slight deterioration (RMSE 88.523 W/kWp, $R^2 = 0.874$). These findings show that carefully designed synthetic features can serve as effective surrogates for unavailable irradiance measurements, preserving the model's ability to reproduce both the smooth profiles observed under clear-sky conditions and the variability induced by transient cloud phenomena.

The systematic comparison presented in this chapter demonstrates that training on actual forecasts reduces the mismatch between the conditions of model development and the inputs encountered in operational use, while synthetic features provide a viable means of mitigating the dependence on proprietary measurements or dedicated instrumentation. The selection and design of meteorological inputs thus emerge as factors as decisive as the choice of learning algorithm itself for the development of photovoltaic forecasting tools that are at once accurate and scalable.

Chapter 4

Optimization model for a VRFB

4.1 Motivations and objectives

The integration of renewable energy sources into modern power systems requires not only accurate forecasting of their production but also the availability of flexible resources capable of mitigating deviations between expected and actual profiles. Forecasts, no matter how accurate, remain affected by intrinsic uncertainty and therefore need to be complemented by technologies that can translate predictive information into reliable operational strategies. In this context, electrochemical storage systems, and in particular BESS, emerge as key enablers for enhancing the value of forecasting and ensuring a more efficient and sustainable energy management. Their ability to rapidly respond, their modularity, and their growing technological maturity make them suitable for providing a broad set of services, including time-shifting of generation, peak shaving, frequency regulation, and the reduction of curtailment [72, 73, 130].

One of the main challenges that still persists is the correct sizing and selection of the most appropriate storage technology. The design of a BESS must carefully consider the nature of the required service, the relevant temporal scale (short-, medium-, or long-term), economic and environmental constraints, as well as the specific technological features of the storage system. Undersizing results in a lack of effectiveness and profitability, while oversizing leads to unjustified costs. In both cases, the economic sustainability of the project is compromised, limiting the long-term viability of the solution.

The scientific literature has extensively investigated optimization models for storage sizing and operation, ranging from linear and mixed-integer programming to stochastic approaches and reinforcement learning [16–18, 131–133]. However, most of these contributions focus predominantly on lithium-ion batteries, neglecting alternative technologies that, although less widespread, exhibit characteristics particularly suitable for long-duration applications. Furthermore, simplified assumptions on efficiency, degradation, or flexibility may lead to misleading results and suboptimal decisions [134, 135].

From these considerations arises the motivation of this chapter: to analyze in a comparative manner two of the most relevant electrochemical storage technologies, lithium-ion batteries and VRFB, by developing an optimization model aimed at determining their optimal sizing in renewable energy integration applications. Lithium-ion batteries hold a leading position in today's energy storage market, driven by declining costs, high efficiency, and mass production largely spurred by the automotive sector [80, 81]. Nevertheless, they are still affected by significant challenges such as vulnerability to thermal runaway, progressive degradation, limited service life, and dependence on critical raw materials, including lithium, nickel, and cobalt[135–140]. These aspects raise concerns about their long-term suitability for grid-scale applications.

On the other hand, vanadium redox flow batteries, although less commercially widespread and associated with higher initial costs, provide distinctive advantages such as the decoupling of power and energy capacity, the possibility of full depth-of-discharge without degradation, long operational lifetimes, and intrinsic safety [83, 84, 141–143]. These characteristics make them particularly attractive for long-duration services, where energy shifting is more relevant than energy density or immediate cost. The growing industrial interest and large-scale deployments, such as the Dalian project in China commissioned in 2022 with an initial capacity of 100 MW/400 MWh [144], confirm their increasing competitiveness in the long-duration storage market.

For both technologies, realistic optimization models are crucial for accurately assessing profitability and system benefits. In the case of lithium-ion batteries, this means explicitly incorporating degradation mechanisms and efficiency losses [88, 145]; for VRFB, it requires representing design flexibility, variable efficiencies, and auxiliary costs associated with electrolyte pumping [93].

In light of these considerations, the objective of the chapter is twofold: first, to develop and apply an optimization model that determines the optimal sizing of a VRFB system integrated with photovoltaic and wind generation; and second, to compare the results with an equivalent lithium-ion system, highlighting differences and trade-offs in terms of annual system costs.

4.1.1 Lithium-ion batteries

Li-ion batteries currently represent the dominant technology in the field of electrochemical storage, thanks to their high performance and the technological maturity achieved after more than thirty years of industrial development. Introduced in the early 1990s and rapidly established in the electric mobility sector, these batteries have also gained a central role in stationary applications, where they are widely employed for frequency regulation, peak shaving, energy arbitrage, and support to the integration of variable renewable energy sources. Their widespread adoption is the result of a combination of favorable characteristics: a high energy density, with values exceeding 200 Wh/kg, a round-trip efficiency typically in the range of 90–95%, very fast response times, and continuously decreasing costs driven by economies of scale in the automotive industry [80, 81].

The operating principle of Li-ion batteries is based on the reversible movement of lithium ions between the two electrodes through the electrolyte. During discharge, lithium ions are released from the anode, generally made of graphite, and migrate towards the cathode, composed of transition metal oxides such as $LiCoO_2$, $LiFePO_4$, or $Li(NiMnCo)O_2$. Electrons, which cannot pass through the electrolyte, flow along the external circuit and provide electric current. The reverse process occurs during charging, when an external power source drives the ions back to the anode. Figure 4.1 shows the simplified structure of a Li-ion cell: it consists of a graphite anode connected to a copper current collector, a lithium-metal oxide cathode supported on an aluminum collector, and a porous separator soaked in electrolyte. The separator allows selective ionic transport while preventing direct contact between the electrodes, thus avoiding short circuits. The combination of these elements enables the reversible intercalation of lithium ions, which constitutes the fundamental storage mechanism.

From a performance perspective, Li-ion batteries are characterized by high coulombic efficiency, which reduces overall system losses, and by good cycle stability. Depending on

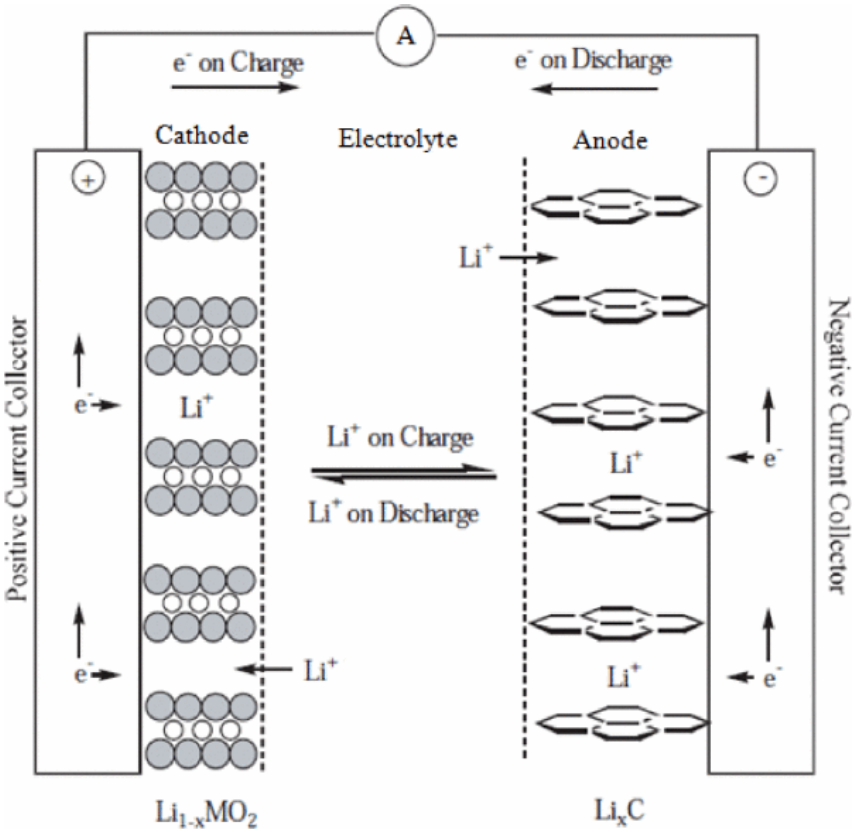


Figure 4.1 – Simplified scheme of a lithium-ion cell. Image retrieved from [146].

the chemistry and operating conditions, the lifetime typically ranges from 3,000 to 7,000 equivalent full cycles, with a calendar life that may exceed ten years. These features explain their suitability for applications requiring rapid response and frequent charge and discharge cycles, such as frequency regulation or the management of daily fluctuations in photovoltaic and wind generation.

Nevertheless, Li-ion batteries also exhibit some critical issues that limit their deployment in certain contexts. Safety is among the most important concerns: the presence of flammable organic electrolytes makes the technology prone to thermal runaway, a potentially dangerous chain reaction triggered by overheating, mechanical damage, or internal short circuits [139]. Another relevant challenge is progressive degradation, caused by mechanisms such as the loss of active lithium, electrode material aging, and the growth of the solid electrolyte interphase. These processes result in capacity fade and increased internal resistance over time [135, 136]. Furthermore, the intrinsic coupling between energy capacity and power output, both determined by the amount of active material in the electrodes, reduces design flexibility and often leads to inefficient oversizing when long discharge durations are required [134].

Finally, Li-ion batteries rely on critical raw materials such as lithium, cobalt, and nickel, whose supply chains are subject to geopolitical risks, price volatility, and environmental and social concerns associated with mining and processing [140].

4.1.2 Vanadium Redox Flow Batteries

Flow batteries represent a class of electrochemical storage systems in which energy is stored in liquid electrolytes kept in external tanks and circulated through an electrochemical cell where the redox reactions take place. Unlike conventional batteries, in which both energy and power are determined by the same cell, flow batteries exhibit a structural decoupling: the *energy capacity* depends exclusively on the volume of electrolytes, while the *power output* is determined by the surface area and the number of cells (stack) [92, 141, 142, 147]. This feature enables significant scalability and design flexibility, allowing the systems to be adapted to diverse application requirements without inefficient oversizing of either the energy or power components. The modern concept of redox flow cells was first formulated by NASA in the 1970s, notably through the pioneering work of Thaller [148] and a series of NASA technical memoranda [149], which laid the foundation for future research.

The first successful demonstration of an all-vanadium configuration was later achieved by Sum and Skyllas-Kazacos in the mid-1980s, who investigated the V(II)/V(III) and V(IV)/V(V) couples [150, 151]. These studies established the basis for VRFBs, which have since evolved into the most technologically advanced type of flow battery for stationary long-duration energy storage. Further electrolyte innovations were introduced in the following decades [152].

Their operating principle relies on reversible redox reactions of vanadium ions in different oxidation states (V^{2+}, V^{3+}, V^{4+}, V^{5+}) dissolved in sulfuric acid solutions. During charging, an external current induces the oxidation of V^{3+} to V^{4+} at the positive electrode and the reduction of V^{4+} to V^{2+} at the negative electrode; the reverse reactions occur during discharge. The two half-cells are separated by a proton exchange membrane that allows H^+ ions to pass while preventing cross-mixing of the electrolytes [83, 141].

A schematic representation of this principle is shown in Figure 4.2. The image illustrates how the electrolytes, containing vanadium ions in different oxidation states, are stored in two external tanks and pumped through the electrochemical stack. A proton exchange membrane separates the two half-cells, ensuring ionic conduction while avoiding electrolyte

crossover. The system is finally connected to the grid via an AC/DC converter that regulates charging and discharging. This scheme highlights the fundamental decoupling between energy and power: while the energy capacity is determined by the electrolyte volume, the power is dictated by the number and size of the stacks [153].

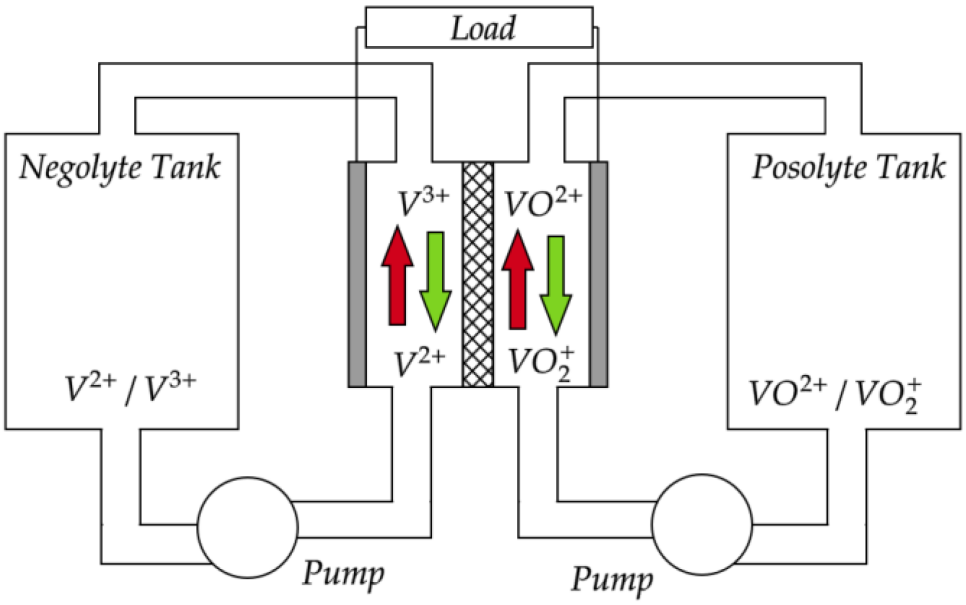


Figure 4.2 – Schematic representation of a vanadium redox flow battery. Image retrieved from [154].

From an operational perspective, VRFBs exhibit several fundamental advantages. First, safety: the use of aqueous, non-flammable electrolytes eliminates the risk of thermal runaway and allows full discharge without permanent damage [84, 143]. Second, long cycle life: since the redox reactions only involve changes in the oxidation state of vanadium ions without structural changes to the electrodes, the system is not subject to the typical degradation phenomena of lithium-ion batteries [141].

The performance of a VRFB is generally evaluated in terms of coulombic efficiency, voltage efficiency, and round-trip efficiency, which can reach 80–85% in laboratory conditions. In real systems, however, auxiliary losses associated with electrolyte pumping must also be considered in techno-economic assessments [147, 155].

In terms of applications, VRFBs are particularly well suited for renewable integration and daily load shifting, thanks to their ability to provide storage times of several hours. In addition, their fast response allows participation in ancillary services for grid balancing and stabilization, although their low energy density limits their use in mobile contexts or where space constraints exist [156, 157].

Despite their advantages, VRFBs still face challenges related to high upfront costs, mainly driven by the price of vanadium and the complexity of auxiliary systems such as pumps and controllers [158]. Nevertheless, large-scale projects are demonstrating the maturity of the technology. The most notable example is the Dalian plant in China, commissioned in 2022 with an initial capacity of 100 MW/400 MWh and planned to be expanded to 200 MW/800 MWh [144].

In the following sections, the integration and optimization of a VRFB system within a hybrid photovoltaic–wind energy context at the Fraunhofer ICT facility will be studied, in order to determine its optimal sizing and long-term economic performance.

4.2 Case study

The problem under investigation concerns the integration of a battery energy storage system in the context of distributed generation and variable demand profiles, with the aim of reducing the total annual system cost. In general, the optimization framework accounts for both the annualized investment cost, calculated through the capital recovery factor (CRF), and the net cost of electricity transactions with the market. For the economic analysis of the system, an operational lifetime of $n = 20$ years is assumed, corresponding to the typical service horizon of vanadium redox flow batteries when deployed in large-scale energy applications[142]. To capture the time value of money in the evaluation of costs and to allow comparability across different investment options, a discount rate of $r = 3\%$ is applied [159]. In the scenario of the VRFB, the optimization determines the optimal storage capacity (kWh) and the installed power (kW, expressed as the number of stacks), in order to identify the techno-economic optimal configuration. In the scenario of the lithium-ion battery, instead, the system size is fixed and set equal to the VRFB optimal size. This choice is intended to enable a fair economic comparison between the two technologies, where the VRFB is optimized in terms of sizing, while the Li-ion battery is evaluated under operational scheduling and degradation scenarios at the same capacity.

The model is implemented in Pyomo, and formulated as a MILP problem, ensuring fast and robust solutions [52]. The model returns the optimal battery configuration in terms of capacity and number of stacks, optimal energy exchanges with the grid, and LCOE. The data used for model development originates from the Fraunhofer ICT Institute in Pfinztal, Germany. The facility comprises a photovoltaic system with a nominal capacity

of about 750 kWp, a wind turbine, and a combined heat and power (CHP) unit with an installed electrical capacity of approximately 370 kW. Production data refer to 2024 and were collected at 15-minute resolution. A schematic representation of the system is provided in Fig. 4.3. For the optimization model, all quantities were subsequently converted into energy by multiplying by the time step $\Delta t = 0.25$ h, ensuring consistency with the energy-based formulation of the problem.

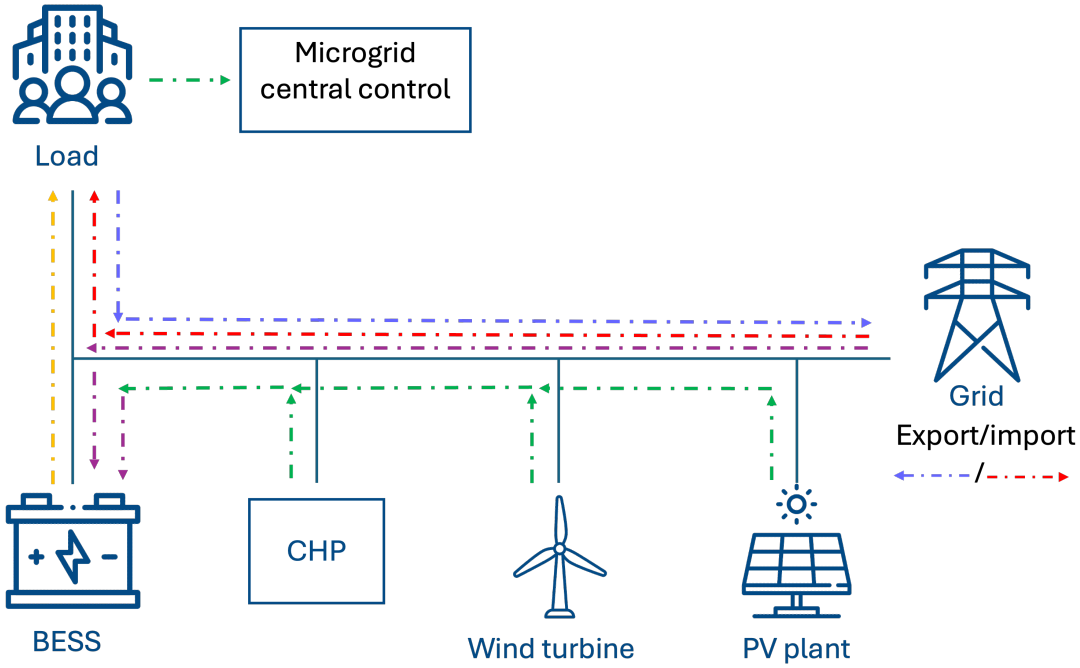


Figure 4.3 – Schematic representation of the microgrid with battery energy storage system (BESS), renewable generation sources, and grid connection.

In addition to generation, the dataset also reports energy exchanges with the grid, namely exported power when production exceeded demand and imported power to cover deficits. On a yearly average, exported power amounted to 49 kW, while imports reached 281 kW. Regarding local generation, the combined output of photovoltaic and wind systems averaged 292 kW, while the CHP unit contributed 116 kW, for a total of 408 kW, with photovoltaics alone accounting for 67 kW. On the demand side, average consumption was 639 kW. These values highlight a significant reliance on the grid, as imports cover a substantial share of the demand. The dataset also includes 2024 German electricity market prices from European Network of Transmission System Operators for Electricity [160]. Since only a single price value is reported for each time step, a margin of 0.15 €/kWh was applied to differentiate

purchase and sale prices, reflecting typical bid-ask spreads in the market. This adjustment allows the model to more realistically simulate energy trading decisions, accounting for the fact that selling electricity usually yields slightly less revenue than the cost of purchasing it.

A few sample rows of the dataset used to build the model are shown in Table 4.1.

4.3 Model setup

4.3.1 Vanadium battery model

The parameters selected for the VRFB include the cost of the electrolyte, the rated power per stack and its cost, and the battery charge and discharge efficiencies. Specifically, the electrolyte cost is set at 100 €/kWh, a value that also includes the cost of the storage tanks required for the system. The stack cost is considered to be 300 €/kW, while the nominal power output per individual stack is fixed at 25 kW [161–163]. The battery is assumed to operate with a charging efficiency of 90% and a discharging efficiency of 75% [164]. All these parameters are summarized in Table 4.2.

The main decision variables in the model are the total installed energy capacity of the VRFB (y) and the number of stacks installed (k), which directly determine the investment cost and the sizing of the system. The total energy capacity defines the maximum amount of energy that can be stored and shifted over time, while the number of stacks sets the system's rated power, i.e., the maximum charging and discharging rates. These two variables capture the fundamental trade-off between storage duration and power availability, which is crucial for both techno-economic analysis and system planning.

Operational variables describe the temporal operation of the system over each time interval $t \in T$, where T denotes the set of discrete time intervals in the optimization horizon and Δt represents the duration of each interval. These variables include the charging energy

Table 4.1 – Example of the rows of the structure of the initial data frame.

Timestamp	Export power (kW)	Import power (kW)	Wind+PV (kW)	CHP (kW)	Consumption (kW)	Price (€/kWh)
2024-02-28 13:00	0	356	234.5	343.9	934.4	0.0580
2024-02-28 13:15	0	408	275.6	345.7	1029.3	0.0604
2024-02-28 13:30	0	272	180.3	358.6	810.9	0.0604
2024-02-28 13:45	0	64	219.6	353.1	636.7	0.0571
2024-02-28 14:00	0	80	216.4	344.2	640.6	0.0487

(x_t^{ch}) and the discharging energy, expressed before and after efficiency losses ($x_t^{\text{dis,raw}}$ and x_t^{dis}), which represent the amounts of energy charged to and discharged from the battery during each interval. The state of charge (s_t) indicates the total stored energy at the end of interval Δt, while q_t^{in} and q_t^{out} denote the energy imported from and exported to the grid, respectively. The electricity demand (f_t) and renewable generation (g_t) are likewise defined as the total energy consumed and produced within each time interval Δt. All quantities are therefore expressed in units of energy, ensuring consistency throughout the formulation. The decision and operational variables are summarized in Table 4.3.

This structure allows the optimization model to capture the interaction among the VRFB, the demand profile, renewable generation, and electricity prices, while enforcing both physical and operational constraints such as charging and discharging efficiencies, state-of-charge limits, and the mutual exclusivity of charging and discharging operations. The optimization is solved with the open-source GLPK solver [165, 166]. This solver, suitable for mixed-integer linear problems of small to intermediate size, guarantees reliable convergence within short computation times.

The optimization problem is formulated as the minimization of the annualized total cost C of the VRFB system. The total cost C, expressed in Eq. 4.2, is given by the sum of two components: (i) the annualized investment cost, which depends on the installed energy capacity and the number of stacks, and (ii) the net operational cost, calculated as the balance between electricity purchased from and sold to the grid over the optimization horizon. The investment cost is annualized through the capital recovery factor (CRF), defined as in Eq. 4.1, where r is the discount rate and n is the economic lifetime of the system:

$$\text{CRF} = \frac{r \cdot (1+r)^n}{(1+r)^n - 1} \tag{4.1}$$

Table 4.2 – Parameters of the VRFB optimization model.

Symbol	Parameter	Unit
c	Electrolyte cost (incl. tanks)	100 €/kWh
e	Stack cost per installed power	300 €/kW
P^{stk}	Rated power of one stack	25 kW
η_{ch}	Charging efficiency	0.90
η_{dis}	Discharging efficiency	0.75

Table 4.3 – Decision and operational variables of the model.

Symbol	Description
k	Number of installed stacks
y	Total installed energy capacity
x_t^{ch}	Charging energy during interval t
$x_t^{\text{dis,raw}}$	Raw discharging energy (before efficiency) during interval t
x_t^{dis}	Effective discharging energy during interval t
s_t	State of charge (SoC, energy stored) at the end of interval t
s_0	Initial state of charge (SoC) of the battery
p_t	Net electricity balance at interval t
q_t^{in}	Energy imported from the grid during interval t
q_t^{out}	Energy exported to the grid during interval t
f_t	Electricity demand (energy) during interval t
g_t	Renewable generation (energy) during interval t

Hence, the objective function is formulated as:

$$\min \left\{ C = \text{CRF} \cdot \left(c \cdot y + e \cdot P^{\text{stk}} \cdot k\right) + \sum_{t \in T} \left(p_t^{\text{buy}} \cdot q_t^{\text{in}} - p_t^{\text{sell}} \cdot q_t^{\text{out}}\right) \right\} \tag{4.2}$$

The optimization problem is subject to a set of constraints that govern both the physical balance of energy flows and the operational boundaries of the system.

As per Eq. (4.3), the state of charge is initialized at 50% of the installed energy capacity:

$$s_0 = 0.5 \cdot y \tag{4.3}$$

This value represents a neutral starting condition that avoids forcing the system to charge or discharge the battery from the outset, allowing optimization to determine the most cost-effective operational schedule of the storage system based on demand, prices, and available generation.

Eq. (4.4) describes the temporal evolution of the stored energy, where the state of charge at each step depends on the value at the previous step and the charged and discharged energy flows:

$$s_t = s_{t-1} + \eta_{\text{ch}} \cdot x_t^{\text{ch}} - x_t^{\text{dis,raw}}, \quad \forall t \in T \tag{4.4}$$

The discharged energy accounts for the discharging efficiency through Eq. (4.5):

$$x_t^{\text{dis}} = \eta_{\text{dis}} \cdot x_t^{\text{dis,raw}}, \quad \forall t \in T \tag{4.5}$$

At each time interval, the electricity demand must be satisfied through renewable generation, battery operation, and grid exchanges, as expressed in Eq. (4.6):

$$g_t + x_t^{\text{dis}} - x_t^{\text{ch}} + q_t^{\text{in}} - q_t^{\text{out}} = f_t, \quad \forall t \in T \tag{4.6}$$

The state of charge is constrained to remain within safe operational limits to prevent deep discharge or overcharge, according to Eq. (4.7):

$$0.05 \cdot y \leq s_t \leq 0.95 \cdot y, \quad \forall t \in T \tag{4.7}$$

The instantaneous energy exchanged by the battery is constrained not to exceed the installed power capacity by means of Eq. (4.8):

$$x_t^{\text{ch}} + x_t^{\text{dis,raw}} \leq P^{\text{stk}} \cdot k \cdot \Delta t, \quad \forall t \in T \tag{4.8}$$

This linear formulation approximates the mutual exclusivity between charging and discharging within each time interval, as binary variables are not introduced to preserve the linearity of the MILP model and ensure compatibility with the GLPK solver.

Finally, Eq. (4.9) defines the design limitations on the maximum number of stacks and the total energy capacity:

$$k \leq 80, \quad y \leq 2000 \text{ kWh} \tag{4.9}$$

Once executed, the algorithm determines the most advantageous configuration of the storage system expressed in terms of installed energy capacity and number of stacks, as well as the temporal scheduling of charging, discharging, and grid exchanges, along with the annualized cost C and the corresponding LCOE [167].

The levelized cost of energy (LCOE) of the VRFB system is computed as:

$$\text{LCOE} = \frac{C}{\sum_{t \in T} f_t} \tag{4.10}$$

where C is the annualized total cost of the system, and the denominator represents the total annual electricity demand satisfied.

4.3.2 Lithium battery model

The optimization model developed for the lithium-ion battery considers exclusively the energy management, while the energy capacity of the battery was fixed at the same size identified as optimal for the vanadium battery. Also, the C-rate of the battery is fixed to maintain the same maximum power output identified for the VRFB. Both charging and discharging efficiencies are set at 90% [168]. The unit cost of the technology was set at 300 €/kWh [169]. All main model parameters are summarized in Table 4.4.

For lithium-ion batteries, two different scenarios are considered, distinguished by the treatment of battery replacement. In both cases, degradation is represented as a linear decrease of usable capacity. Although this assumption is a simplification, it is widely adopted in techno-economic analyses due to its tractability and the lack of detailed operational data, while more sophisticated ageing models usually require extensive calibration and introduce significant computational complexity [170, 171]. A linear formulation was therefore chosen as a conservative assumption in favour of lithium-ion technology, acknowledging that non-linear ageing mechanisms could lead to faster performance losses in practice. The decision variables maintained in the model concern the energy charged into and discharged from the battery at each time step, the state of charge, the energy imported from the grid, and the energy exported to the grid. The constraints applied are essentially the same as those defined for the vanadium battery, except that the rated power is fixed and the usable capacity decreases according to the degradation model.

In the no-replacement scenario, the battery is operated for the full 20-year lifetime of the plant without replacement. The capacity decreases linearly by 57% over the period, corresponding to an annual degradation rate of approximately 2.8%. The cost evaluation in

Table 4.4 – Parameters of the Lithium-ion optimization model.

Symbol	Parameter	Unit
c	Total cost	300 €/kWh
η_{ch}	Charging efficiency	0.90
η_{dis}	Discharging efficiency	0.90

this case includes only the initial investment and the net balance of electricity exchanges with the grid, which are directly influenced by the progressive reduction in usable capacity.

In the replacement scenario, a complete replacement is introduced after 14 years of operation. This value falls within the 10–15 year lifetime range typically reported for utility-scale lithium-ion systems [172], and it is set slightly before the upper bound of this range to ensure a prudent assumption consistent with established benchmarks. In the first cycle, the capacity decreases linearly by 40% (2.8%/year). After replacement, the capacity is restored to its nominal value and then decreases by a further 17% over the remaining 6 years. The total present cost is calculated as the sum of the initial investment and the discounted replacement cost, while the equivalent annual cost is derived by applying the CRF (Eq. (4.1)) to this value.

4.4 Results

This section presents the main outcomes of the analysis and the results of the optimization. The techno-economic comparison is carried out for four scenarios:

- absence of storage (i.e., current condition),
- use of a vanadium redox flow battery (VRFB),
- use of a lithium-ion battery, including degradation and replacement after 14 years,
- use of a lithium-ion battery, considering only degradation over 20 years.

The key performance indicators i.e., storage capacity, rated power, LCOE, total annualized cost, and grid exchanges obtained from the optimal configurations are summarized in Table 4.5. The optimal sizing indicates that the VRFB system requires twenty-four stacks, with a rated power of 600 kW and an energy of 2 MWh. For the lithium-ion battery, the nominal was therefore fixed at 2000 kWh, the same size identified as optimal for the VRFB, and the maximum power output was set to 600 kW by selecting a C-rate of 0.3 [168]. When annual degradation is considered, the lithium-ion system shows a residual capacity of 1.67 MWh, corresponding to an available power of 500.7 kW. This reflects the reduction in usable energy capacity over time due to degradation, whereas the VRFB maintains its nominal performance throughout its lifetime. Overall, both storage technologies significantly

improve the system economics compared to the no-storage case, although their performance varies depending on the assumptions related to degradation and replacement.

Table 4.5 – Comparison of techno-economic results for the different scenarios.

Metric	No Storage	VRFB	Li-Ion (replacement)	Li-Ion (no replacement)
Total annual cost (k€)	358	333	384	348
Residual capacity (kWh)	–	2000	1669	1669
Installed power (kW)	–	600	500.7	500.7
LCOE (€/kWh)	0.092	0.085	0.098	0.089
Energy imported from grid (GWh)	2.04	1.86	1.85	1.85
Energy exported to grid (GWh)	1.73	1.33	1.39	1.39

Without any storage system, the total annualized cost of the microgrid amounts to approximately 358 k€. Integrating a vanadium redox flow battery reduces this value to around 333 k€, corresponding to a reduction of about 7% compared to the current configuration. The VRFB thus represents the most economically advantageous solution, thanks to its stable performance over time and negligible degradation and replacement costs.

Among the lithium-ion configurations, the case that does not consider substitution reaches a total annualized cost of 348 k€, making it the second most favorable option, with performance comparable to the VRFB. In contrast, when battery replacement after 14 years is included, the total cost rises to 384 k€, becoming the worst-case scenario, even more expensive than the no-storage configuration. This highlights the strong economic impact of periodic battery replacement, which can offset the operational advantages of lithium-ion technology in long-term stationary applications.

In terms of the LCOE, the VRFB achieves the lowest value (0.085 €/kWh), followed by the lithium-ion system considering only degradation (0.089 €/kWh). The replacement-based lithium-ion configuration shows the highest value (0.098 €/kWh), confirming its lower cost-effectiveness. The no-storage scenario remains at an intermediate level (0.092 €/kWh). Grid exchanges are also considerably reduced by the integration of storage: compared with 2.04 GWh imported and 1.73 GWh exported without storage, the VRFB limits these exchanges to 1.86 GWh and 1.33 GWh, respectively. The lithium-ion systems exhibit similar behavior, with about 1.85 GWh imported and 1.39 GWh exported, consistent with their higher round-trip efficiency.

Figures 4.4–4.5 illustrate the hourly energy flows on a representative spring day for the storage configurations. In all cases, the curves represent demand (black), renewable genera-

tion (green), energy imported from the grid (blue), exported energy (red), battery charging energy (orange), and battery discharging energy (purple).

In the VRFB scenario (Figure 4.4), the battery shows a smooth charge–discharge pattern: it absorbs part of the midday surplus and releases energy gradually in the evening, effectively reducing grid imports. This results in stable and balanced operation, confirming the suitability of flow batteries for long-duration energy shifting.

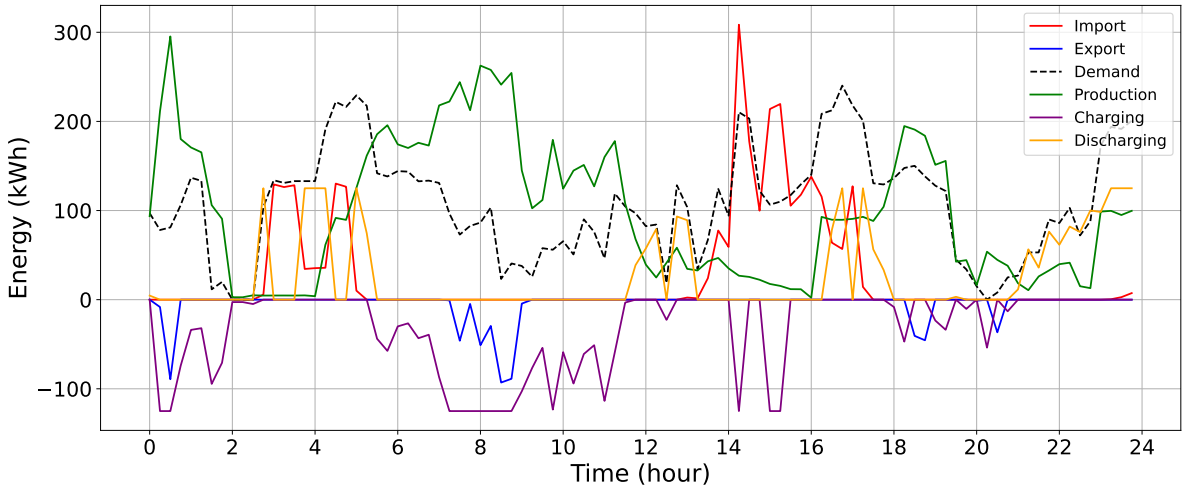


Figure 4.4 – Hourly energy flows for the VRFB scenario.

Since both lithium-ion configurations (with and without replacement) exhibit the same operational behavior, a single representative profile is shown in Figure 4.5. The difference between the two cases lies in the economic treatment of degradation and replacement, which affects the effective capacity of the battery after year 14, but does not alter the short-term operational patterns.

In the lithium-ion scenario, the battery exhibits faster and more frequent charge–discharge cycles compared to the VRFB. Most of the midday surplus is absorbed, reducing exports, while stored energy is released in the evening to lower grid imports. The operation is characterized by high round-trip efficiency and rapid response, allowing effective short-term energy shifting.

Both storage options enable a more balanced and predictable interaction with the grid. The VRFB ensures smoother operation with lower variability, while lithium-ion batteries excel in fast response and high-efficiency cycling. Although the lithium-ion system without replacement achieves comparable economic performance, the VRFB stands out as a more

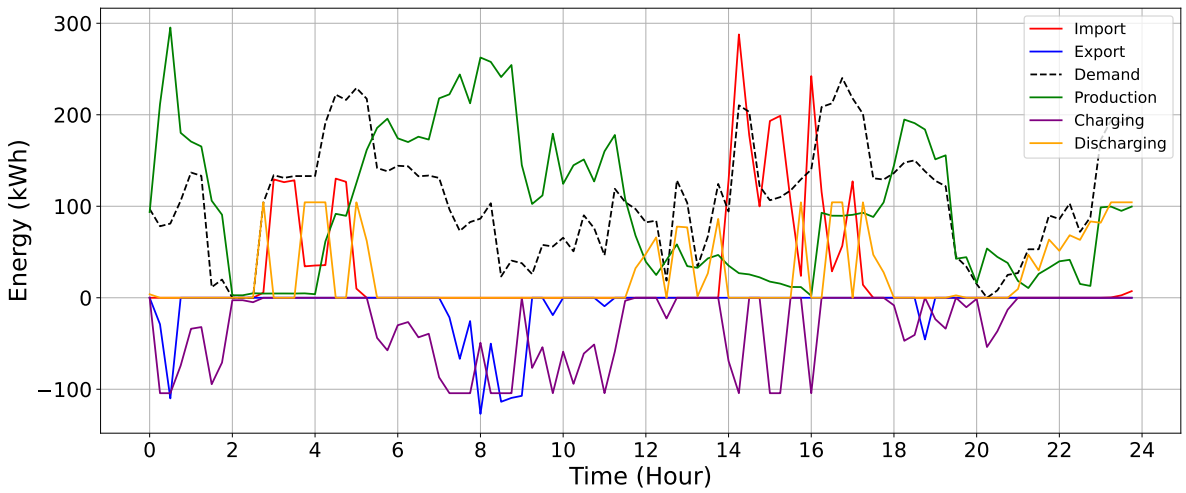


Figure 4.5 – Hourly energy flows for the lithium-ion scenarios (with and without replacement).

robust long-term solution due to its negligible degradation, high reliability, and the intrinsic safety of its non-flammable vanadium electrolyte.

A further advantage of the VRFB is its modular design: the energy capacity can be increased simply by adding electrolyte, without modifying the installed power. This makes it easily adaptable to future photovoltaic expansions or to the integration of additional renewable sources, ensuring scalability and lower upgrade costs. In contrast, lithium-ion systems feature a more rigid configuration, where energy and power are coupled within fixed battery modules, limiting their scalability potential.

Therefore, in the analyzed scenario, VRFBs prove to be not only economically competitive but also a sustainable and durable solution for large-scale stationary storage, offering safety, robustness, and flexibility for future energy system development.

4.5 Chapter summary

This chapter shifted the focus from forecast accuracy to the operational means of enhancing their value through storage, analyzing how technology selection and sizing decisions affect the techno-economic value of integrating batteries into hybrid renewable systems. While lithium-ion batteries dominate today's stationary storage market, their intrinsic coupling between energy and power, safety constraints, and degradation dynamics motivate comparison with VRFBs. The latter, with their decoupled architecture and long service

life, are particularly suited for multi-hour energy shifting. To ensure a meaningful operational comparison, an optimization model was developed that co-determines capacity and power output for VRFBs and evaluates lithium-ion alternatives under consistent operating assumptions.

The modeling framework was formulated as a mixed-integer linear programming problem implemented in Pyomo, with the objective of minimizing the total annualized cost over a 20-year horizon. Investment costs were annualized using the capital recovery factor, while operational costs reflected the balance of electricity transactions with the market under time-varying prices. The model enforced energy balance, round-trip efficiencies, power limits, and state-of-charge constraints, while incorporating the technological features of each system: independent scalability of energy and power for VRFBs, versus fixed C-rate operation and capacity fade for lithium-ion batteries. For the latter, two scenarios were considered to represent realistic asset management strategies: operation with linear degradation only, and operation with linear degradation plus full replacement after year fourteen. The assumption of linear capacity degradation for lithium-ion batteries is a simplifying but widely adopted choice in long-term techno-economic analyses. While real aging mechanisms are nonlinear and depend on operating conditions, a linear approximation enables a transparent and computationally efficient comparison of different asset management strategies, such as continued operation with reduced capacity versus mid-life replacement. Consequently, the results should be interpreted as capturing relative trends and strategic trade-offs rather than providing precise lifetime or cost predictions for lithium-ion systems.

Using high-resolution data from 2024 at the Fraunhofer ICT microgrid (including photovoltaics, wind, combined heat and power, and grid exchanges), the optimization identified for the VRFB an optimal configuration of 2 MWh and 600 kW (twenty-four stacks), which was then adopted as the reference size for lithium-ion scenarios. Compared with the no-storage case (total annualized cost of about 358 k€), the VRFB reduced costs to approximately 333 k€, achieved the lowest LCOE (0.085 €/kWh), and decreased both grid imports (to 1.86 GWh) and exports (to 1.33 GWh). The lithium-ion system with degradation but no replacement reached a comparable LCOE (0.089 €/kWh) and an annualized cost of about 348 k€, while including a full replacement raised the cost to around 384 k€ with an LCOE of 0.098 €/kWh. Operationally, the VRFB provided smoother multi-hour shifting aligned with daily renewable variability, whereas lithium-ion batteries favored faster and more frequent cycles with higher round-trip efficiency.

Overall, the results suggest that in long-duration applications with significant daily shifting needs, VRFBs can deliver robust economic performance while preserving nominal capacity throughout their service life and offering modular scalability of energy and power. Lithium-ion solutions remain competitive when high cycle efficiency and fast response are primary requirements, but their long-term cost-effectiveness is sensitive to assumptions on degradation and replacement. The findings underscore that storage technology selection and sizing strategy are as decisive as forecasting methods when the goal is to translate predictive information into durable system value. Future extensions may incorporate stochastic price scenarios, forecast uncertainty, more detailed aging models, and multi-service co-optimization to further test technology choices under realistic operating conditions.

Chapter 5

Conclusions and Future Research

5.1 Summary and Contributions

This thesis has addressed two of the most critical challenges in the integration of renewable energy into modern power systems: the forecasting of variable generation and the optimal deployment of battery energy storage systems. By combining methodological advances in photovoltaic power forecasting with a comparative techno-economic analysis of lithium-ion and vanadium redox flow batteries, the work provides both scientific insights and practical contributions to the development of resilient and cost-effective energy infrastructures.

The main original contributions of the thesis can be summarized as follows:

- A comparative analysis of photovoltaic power forecasting models trained on local ground-based and satellite-derived meteorological datasets, highlighting the trade-off between forecasting accuracy, data accessibility, robustness, and scalability across different locations and operating conditions.
- The development and validation of forecast-based training approaches that directly exploit numerical weather prediction data, demonstrating improved operational accuracy compared to models relying solely on historical measurements.
- The introduction of synthetic features for photovoltaic power forecasting, showing that reliable model performance can be preserved even in the absence of direct irradiance measurements, thus enhancing the scalability and applicability of forecasting systems in data-scarce environments.

- The formulation and application of an optimization model for electrochemical energy storage systems, enabling a detailed techno-economic comparison between lithium-ion batteries and vanadium redox flow batteries, including efficiency, degradation, lifetime, auxiliary losses, and cost aspects. These results have been derived from the work presented in [121].
- The integrated assessment of forecasting accuracy and storage system sizing, demonstrating that predictive performance and storage deployment must be jointly considered to fully exploit the benefits of renewable energy integration.

A further original contribution of the thesis lies in the integration of photovoltaic power forecasting and energy storage optimization within a unified framework. The results indicate that accurate forecasting alone is insufficient if not paired with appropriately dimensioned storage systems, and conversely, that oversized or inadequately managed storage cannot fully leverage the benefits of advanced predictive models. This finding emphasizes the need for a holistic approach to the design of future energy systems, where forecasting and storage technologies are co-optimized to achieve cost-effectiveness, stability, and sustainability.

In addition to its methodological contributions, the thesis provides empirical validation through a real-world case study based on operational data from the Fraunhofer ICT facility. By applying the developed models to real operational data, the research validates its findings under practical conditions, demonstrating the robustness of the methods and their potential for real-world applications. The case study also illustrates how the choice of storage technology impacts long-term system economics, with vanadium redox flow batteries emerging as a competitive and durable solution for long-duration services, despite their higher initial costs.

From a broader perspective, the thesis contributes to the discourse on energy policy and technological strategy. The findings underscore the importance of technological diversification: while lithium-ion batteries remain highly effective for short-term, high-power applications, vanadium redox flow batteries offer distinctive advantages for long-duration, grid-scale storage. The work therefore suggests that future energy planning should avoid an exclusive reliance on a single technology and instead promote a balanced portfolio, leveraging the strengths of complementary storage solutions to address a variety of services and time horizons.

5.2 Proposal for Future Research

Building on the findings and contributions of this thesis, several promising directions for future research can be identified. These directions reflect both the methodological limitations encountered in the present work and the broader challenges faced by the energy sector in its transition towards large-scale renewable integration.

A first area for further research concerns the enhancement of photovoltaic power forecasting models. While the present study has demonstrated the effectiveness of synthetic features and forecast-based training, future work could extend this framework by incorporating hybrid data sources. For instance, the integration of satellite imagery, sky cameras, and high-resolution remote sensing could further refine cloud representation and irradiance estimation. Similarly, probabilistic forecasting methods should be developed to explicitly quantify uncertainty, enabling system operators to design risk-aware strategies that account for the stochastic nature of weather-driven generation. The exploration of deep learning architectures, such as recurrent neural networks and transformers, could also provide additional gains in predictive accuracy and temporal resolution.

A second promising direction involves the improvement of battery optimization models. In this thesis, the degradation of lithium-ion systems has been represented using linear approximations, while auxiliary losses in vanadium redox flow batteries have been treated in a simplified manner. Future research should incorporate more detailed ageing models, reflecting non-linear and path-dependent degradation mechanisms that depend on temperature, cycling patterns, and depth of discharge. For VRFBs, the representation of electrolyte crossover, state-of-charge imbalances, and capacity fade over extended operation should be refined. Moreover, multi-objective optimization frameworks could be developed to simultaneously capture technical, economic, and environmental dimensions, thus offering a more comprehensive evaluation of storage investments.

A third dimension concerns the market and policy context in which storage systems operate. The present work has assumed a simplified energy trading framework; however, real-world storage deployment increasingly interacts with complex electricity markets that include capacity payments, ancillary service revenues, and dynamic tariff schemes. Future research should therefore integrate detailed market models, exploring how storage systems can be optimally scheduled under varying price signals and regulatory environments. The

inclusion of carbon pricing and environmental externalities would also provide a more complete picture of the long-term societal value of different storage technologies.

Another important research direction is the exploration of hybrid storage configurations. Rather than evaluating lithium-ion and VRFB systems in isolation, future studies should consider their joint deployment, where the high power density and fast response of lithium-ion batteries could complement the long-duration capabilities and stability of VRFBs. Such hybrid architectures could yield superior performance in terms of flexibility, resilience, and cost-effectiveness, particularly in microgrid and islanded system contexts.

Finally, scaling up the analysis to larger systems and different geographical locations represents a crucial step. The present case study, based on the Fraunhofer ICT facility, provided valuable empirical validation, but future research should apply the models to diverse climatic zones, market structures, and grid topologies. Large-scale simulations that account for regional renewable variability, transmission constraints, and interconnection opportunities would allow researchers to generalize the findings and support policy decisions at national or continental levels.

In summary, future research should move towards more realistic, integrated, and context-specific analyses of renewable forecasting and storage. Methodological advancements in data assimilation, machine learning, and optimization must be combined with broader considerations of markets, policies, and sustainability. By pursuing these directions, the academic and industrial communities can contribute to the design of next-generation energy systems that are not only technically efficient but also economically viable, environmentally responsible, and socially beneficial.

Bibliography

[1] International Energy Agency. *Renewables 2023 – Executive Summary*. https://www.iea.org/reports/renewables-2023/executive-summary. Accessed: 2025-10-17. 2023.

[2] D. Gielen et al. 'The role of renewable energy in the global energy transformation'. In: *Energy Strategy Reviews* 24 (2019), pp. 38–50.

[3] Claire Dupont, Sebastian Oberthür and Ingmar von Homeyer. 'Towards a Climate-Neutral Union by 2050? The European Green Deal, Climate Law and Green Recovery'. In: *Politics and Governance* 10.1 (2022), pp. 5–15. DOI: 10.17645/pag.v10i1.4910.

[4] Christian Egenhofer, Jorge Núñez Ferrer and Fabio Genoese. 'Europe's Climate Target for 2050: An Assessment'. In: *Intereconomics* 56.6 (2021), pp. 366–372. DOI: 10.1007/s10272-021-0999-0.

[5] Cesare Mallia, George Papaefthymiou and Pantelis Capros. 'Examining pathways for a climate neutral Europe by 2050'. In: *Energy* 310 (2025), p. 130823. DOI: 10.1016/j.energy.2025.130823.

[6] International Energy Agency. *Renewables 2023 – Electricity*. https://www.iea.org/reports/renewables-2023/electricity. Accessed: 2025-10-17. 2023.

[7] Md Shafiullah, Shakir D. Ahmed and Fahad A. Al-Sulaiman. 'Grid Integration Challenges and Solution Strategies for Solar PV Systems: A Review'. In: *IEEE Access* 10 (2022), pp. 52233–52257. DOI: 10.1109/ACCESS.2022.3174555.

[8] Gowthamraj Rajendran, Reiko Raute and Cedric Caruana. 'A Comprehensive Review of Solar PV Integration with Smart-Grids: Challenges, Standards, and Grid Codes'. In: *Energies* 18.9 (2025), p. 2221. DOI: 10.3390/en18092221.

[9] P. Di Leo et al. 'Advancements and Challenges in Photovoltaic Power Forecasting: A Comprehensive Review'. In: *Energies* 18.8 (2025), p. 2108. DOI: 10.3390/en18082108.

[10] Marta Lafuente-Cacho, Óscar Izquierdo-Monge and Paula Peña. 'State of the Art for Solar and Wind Energy-Forecasting Methods for Sustainable Grid Integration'. In: *Journal of Power Electronics and Power Systems* 12.13 (2025). DOI: 10.1007/s40518-025-00262-z.

[11] M. Hosenuzzaman et al. 'Global prospects, progress, policies, and environmental impact of solar Photovoltaic power generation'. In: *Renewable and Sustainable Energy Reviews* 41 (2015), pp. 284–297.

[12] S. K. Firth, K. J. Lomas and S. J. Rees. 'A simple model of PV system performance and its use in fault detection'. In: *Solar Energy* 84.4 (2010), pp. 624–635.

[13] L. M. Ayompe et al. 'Validated real-time energy models for small-scale grid-connected PV-systems'. In: *Energy* 35.10 (2010), pp. 4086–4091.

[14] A. Rahman et al. 'Environmental impact of renewable energy source based electrical power plants: Solar, wind, hydroelectric, biomass, geothermal, tidal, ocean, and osmotic'. In: *Renewable and Sustainable Energy Reviews* 161 (2022), p. 112279.

[15] B. Taghezouit et al. 'Intelligent Monitoring of Photovoltaic Systems via Simplicial Empirical Models and Performance Loss Rate Evaluation under LabVIEW: A Case Study'. In: *Energies* 15.21 (2022).

[16] Naoto Kakimoto and Ryo Asano. 'Linear operation of photovoltaic array with directly connected lithium-ion batteries'. In: *IEEE Transactions on Sustainable Energy* 8.4 (2017), pp. 1647–1657.

[17] Holger C Hesse et al. 'Lithium-ion battery storage for the grid—A review of stationary battery storage system design tailored for applications in modern power grids'. In: *Energies* 10.12 (2017), p. 2107.

[18] Xing Luo et al. 'Overview of current development in electrical energy storage technologies and the application potential in power system operation'. In: *Applied energy* 137 (2015), pp. 511–536.

[19] International Energy Agency. *Batteries and Secure Energy Transitions*. https://www.iea.org/reports/batteries-and-secure-energy-transitions/status-of-battery-demand-and-supply. Accessed: 2025-08-19. 2024.

[20] J. Antonanzas et al. 'Photovoltaic power forecasting'. In: *Solar Energy* 136 (2016), pp. 78–111.

[21] Muhammad Qamar Raza and Abbas Khosravi. 'A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings'. In: *Renewable and Sustainable Energy Reviews* 50 (2015), pp. 1352–1372.

[22] Cyril Voyant et al. 'Machine learning methods for solar radiation forecasting: A review'. In: *Renewable Energy* 105 (2017), pp. 569–582. DOI: 10.1016/j.renene.2016.12.095. URL: https://doi.org/10.1016/j.renene.2016.12.095.

[23] Elizaveta Kharlova, Daniel May and Petr Musilek. 'Forecasting photovoltaic power production using a deep learning sequence to sequence model with attention'. In: *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2020, pp. 1–7.

[24] Yu Li et al. 'Very short-term power prediction for PV systems based on stochastic time series models'. In: *Applied Energy* 96 (2012), pp. 378–386. DOI: 10.1016/j.apenergy.2011.11.050.

[25] Leticia de Oliveira Santos et al. 'Photovoltaic power estimation and forecast models integrating physics and machine learning: A review on hybrid techniques'. In: *Solar Energy* 284 (2024), p. 113044.

[26] N Martin and JM Ruiz. 'Calculation of the PV modules angular losses under field conditions by means of an analytical model'. In: *Solar energy materials and solar cells* 70.1 (2001), pp. 25–38.

[27] Riccardo Adinolfi Borea et al. 'Impact of environmental variables on tilt selection for energy yield maximization in bifacial photovoltaic modules: Modeling review and parametric analysis'. In: *Applied Sciences* 14.24 (2024), p. 11497.

[28] Lina M Shaker et al. 'Examining the influence of thermal effects on solar cells: a comprehensive review'. In: *Sustainable Energy Research* 11.1 (2024), p. 6.

[29] Salam J Yaqoob et al. 'Comparative study with practical validation of photovoltaic monocrystalline module for single and double diode models'. In: *Scientific reports* 11.1 (2021), p. 19153.

[30] Rodolfo Manuel Arias García and Ignacio Pérez Abril. 'Photovoltaic module model determination by using the Tellegen's theorem'. In: *Renewable Energy* 152 (2020), pp. 409–420.

[31] Chih-Hao Chang, Jia-Jun Zhu and Huan-Liang Tsai. 'Model-based performance diagnosis for PV systems'. In: *Proceedings of SICE Annual Conference 2010*. 2010, pp. 2139–2145.

[32] M. J. Mayer and G. Gróf. 'Extensive comparison of physical models for photovoltaic power forecasting'. In: *Applied Energy* 283 (2021), p. 116267. DOI: 10.1016/j.apenergy.2020.116267.

[33] A. Dolara, S. Leva and G. Manzolini. 'Comparison of different physical models for PV power output prediction'. In: *Renewable Energy* 80 (2015), pp. 243–252. DOI: 10.1016/j.renene.2015.01.060.

[34] E. Ogliari et al. 'Physical and hybrid methods comparison for the day-ahead PV output power forecast'. In: *Renewable Energy* 113 (2017), pp. 11–21. DOI: 10.1016/j.renene.2017.05.079.

[35] Muhammad Qamar Raza, Mithulananthan Nadarajah and Chandima Ekanayake. 'On recent advances in PV output power forecast'. In: *Solar Energy* 136 (2016), pp. 125–144.

[36] Diego Delgado and Osvaldo A. Rosso. 'A review of solar radiation forecasting methods'. In: *Energy Reports* 5 (2019), pp. 829–846. DOI: 10.1016/j.egyr.2019.06.010.

[37] Maria Grazia De Giorgi, Paolo Maria Congedo and Maria Malvoni. 'Photovoltaic power forecasting using statistical methods: impact of weather data'. In: *IET Science, Measurement & Technology* 8.3 (2014), pp. 90–97.

[38] Min Sun et al. 'Comparison of statistical and machine learning methods for solar radiation forecasting'. In: *Journal of Cleaner Production* 271 (2020), p. 122129. DOI: 10.1016/j.jclepro.2020.122129.

[39] Rob J. Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. 2nd. online edition, freely available at otexts.com/fpp2. OTexts, 2018.

[40] Maxime Taillardat et al. 'Calibrated Ensemble Forecasts Using Quantile Regression Forests and Ensemble Model Output Statistics'. In: *Monthly Weather Review* 144.6 (2016), pp. 2375–2393.

[41] J.M. Ruiz and et al. 'Forecasting solar photovoltaic power production: A comprehensive review'. In: *Energies* 17.16 (2024), p. 4145. DOI: 10.3390/en17164145.

[42] Ioannis P. Panapakidis and Georgios C. Christoforidis. 'A hybrid ANN/GA/ANFIS model for very short-term PV power forecasting'. In: *2017 11th IEEE International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG)*. IEEE. 2017, pp. 1–6.

[43] Fatemeh Raeisi et al. 'Machine Learning Based Solar Photovoltaic Power Forecasting: A Review and Comparison'. In: *Renewable and Sustainable Energy Reviews* 133 (2020), p. 110300. DOI: 10.1016/j.rser.2020.110300.

[44] Yi Zhang and et al. 'A review of PV power forecasting using machine learning techniques'. In: *Energy Reports* 9 (2023), pp. 2950–2965. DOI: 10.1016/j.egyr.2023.04.010.

[45] J. Shi et al. 'Forecasting Power Output of Photovoltaic Systems Based on Weather Classification and Support Vector Machines'. In: *IEEE Transactions on Industry Applications* 48 (2012), pp. 1064–1069.

[46] J.M. Ruiz, R. Perez and et al. 'Deep learning models for PV power forecasting: Review'. In: *Energies* 17.16 (2024), p. 3973. DOI: 10.3390/en17163973.

[47] M. Schmid and T. Hothorn. 'Flexible boosting of accelerated failure time models'. In: *BMC Bioinformatics* 9 (2008), pp. 1–13.

[48] Razin Ahmed et al. 'A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization'. In: *Renewable and Sustainable Energy Reviews* 124 (2020), p. 109792.

[49] Manoja Kumar Behera et al. 'Solar photovoltaic power forecasting using optimized modified extreme learning machine technique'. In: *Engineering Science and Technology, an International Journal* 21.3 (2018), pp. 428–438.

[50] Ajay Kumar Bansal. 'A review of recent advances in solar PV power forecasting techniques'. In: *Sustainability* 14.24 (2022), p. 17005. DOI: 10.3390/su142417005.

[51] Utpal Kumar Das et al. 'AI explainability and governance in smart energy systems: A review'. In: *Frontiers in Energy Research* 11 (2023), p. 1071291. DOI: 10.3389/fenrg.2023.1071291.

[52] Chunxia Gao, Zhaoyan Zhang and Peiguang Wang. 'Day-ahead scheduling strategy optimization of electric–thermal integrated energy system to improve the proportion of new energy'. In: *Energies* 16.9 (2023), p. 3781.

[53] Abd Alhamid Rabia Khattab, Nada Mohamed Elshennawy and Mahmoud Fahmy. 'GMA: gap imputing algorithm for time series missing values'. In: *Journal of Electrical Systems and Information Technology* 10.1 (2023), p. 41.

[54] R. Marquez et al. 'Forecasting of Global Horizontal Irradiance Using Sky Cover Indices'. In: *Journal of Solar Energy Engineering* 135.1 (2013), p. 011017.

[55] Jie Zhang et al. 'Machine Learning Based Solar Photovoltaic Power Forecasting for Distinct Climatic Regions'. In: *Energy Science & Engineering* 11.5 (2023), p. 70013. DOI: 10.1002/ese3.70013.

[56] Adel Mellit and Saifur Rahman Kalogirou. 'Artificial intelligence techniques for photovoltaic applications: A review'. In: *Progress in Energy and Combustion Science* 34.5 (2018), pp. 699–726. DOI: 10.1016/j.pecs.2018.04.001.

[57] Utpal Kumar Das et al. 'AI explainability and governance in smart energy systems: A review'. In: *Frontiers in Energy Research* 11 (2023), p. 1071291. DOI: 10.3389/fenrg.2023.1071291.

[58] Martin János Mayer. 'Benefits of physical and machine learning hybridization for photovoltaic power forecasting'. In: *Renewable and Sustainable Energy Reviews* 168 (2022), p. 112772.

[59] Adel Mellit and Soteris A. Kalogirou. 'Artificial intelligence techniques for photovoltaic applications: A review'. In: *Renewable and Sustainable Energy Reviews* 13.2 (2009), pp. 478–500.

[60] Iza Sazanita Isa et al. 'Weather forecasting using photovoltaic system and neural network'. In: *2010 2nd International Conference on Computational Intelligence, Communication Systems and Networks*. IEEE. 2010, pp. 1–5.

[61] First Author and Second Author. 'Optimizing photovoltaic power plant forecasting with hyperparameter optimization of FF-ANN and GA techniques'. In: *Scientific Reports* 14.1 (2024), p. 12345.

[62] First Author and Second Author. 'PV Production Forecast Using Hybrid Models of Time Series with AI tools'. In: *Energies* 18.11 (2025), p. 2692.

[63] Kasun Amarasinghe, Daniel Marino and Milos Manic. 'Ensemble models for solar power forecasting: A weather classification approach'. In: *AIMS Energy* 8.2 (2020), pp. 252–271.

[64] Ijeoma Iheanetu et al. 'A comprehensive review on ensemble solar power forecasting'. In: *Journal of Electrical Engineering & Technology* (2023).

[65] First Author and Second Author. 'Dynamic model selection in a hybrid ensemble framework for robust photovoltaic power forecasting'. In: *Applied Energy* 350 (2024), p. 122345.

[66] Jinwook Rhyu et al. 'Automated outlier detection and estimation of missing data'. In: *Computers & Chemical Engineering* 180 (2024), p. 108448.

[67] Yang Zhang, Nirvana Meratnia and Paul Havinga. 'Outlier detection techniques for wireless sensor networks: A survey'. In: *IEEE communications surveys & tutorials* 12.2 (2010), pp. 159–170.

[68] M. S. Hossain and H. Mahmood. 'Short-term photovoltaic power forecasting using an LSTM neural network and synthetic weather forecast'. In: *IEEE Access* 8 (2020), pp. 172524–172533.

[69] J. Wang et al. 'Exploring Key Weather Factors From Analytical Modeling Toward Improved Solar Power Forecasting'. In: *IEEE Transactions on Smart Grid* 10.2 (Mar. 2019), pp. 1417–1427.

[70] I. A. Abdelmoula, S. Elhamaoui, O. Elalani et al. 'A photovoltaic power prediction approach enhanced by feature engineering and stacked machine learning model'. In: *Energy Reports* 8 (2022), pp. 1288–1300.

[71] Y. Ledmaoui et al. 'Enhanced Fault Detection in Photovoltaic Panels Using CNN-Based Classification with PyQt5 Implementation'. In: *Sensors* 24 (2024), p. 7407.

[72] Matthew T. Lawder et al. 'Battery Energy Storage System (BESS) and Battery Management System (BMS) for Grid-Scale Applications'. In: *Proceedings of the IEEE* 102.6 (2014), pp. 1014–1030. DOI: 10.1109/JPROC.2014.2317451.

[73] Héricles Eduardo Oliveira Farias and Luciane Neves Canha. 'Battery energy storage systems (BESS) overview of key market technologies'. In: *2018 IEEE PES Transmission & Distribution Conference and Exhibition-Latin America (T&D-LA)*. IEEE. 2018, pp. 1–5.

[74] Georgios Karmiris and Tomas Tengnér. 'Peak shaving control method for energy storage'. In: *Corporate Research Center, Vasterås, Sweden* (2013).

[75] Yuanyuan Shi et al. 'Using battery storage for peak shaving and frequency regulation: Joint optimization for superlinear gains'. In: *IEEE transactions on power systems* 33.3 (2017), pp. 2882–2894.

[76] Ujjwal Datta, Akhtar Kalam and Juan Shi. 'A review of key functionalities of battery energy storage system in renewable energy integrated power systems'. In: *Energy Storage* 3.5 (2021), e224.

[77] Yue Zhang and Anurag Srivastava. 'Voltage control strategy for energy storage system in sustainable distribution system operation'. In: *Energies* 14.4 (2021), p. 832.

[78] Paul Denholm et al. 'The four phases of storage deployment: A framework for the expanding role of storage in the US power system'. In: *Golden, CO* (2021).

[79] Abraham Alem Kebede et al. 'A comprehensive review of stationary energy storage devices for large scale renewable energy sources grid integration'. In: *Renewable and sustainable energy reviews* 159 (2022), p. 112213.

[80] Micah S Ziegler, Juhyun Song and Jessika E Trancik. 'Determinants of lithium-ion battery technology cost decline'. In: *Energy & Environmental Science* 14.12 (2021), pp. 6074–6098.

[81] Björn Nykvist and Måns Nilsson. 'Rapidly falling costs of battery packs for electric vehicles'. In: *Nature Climate Change* 5.4 (2015), pp. 329–332.

[82] Patrick T Moseley and David AJ Rand. 'High-temperature sodium batteries for energy storage'. In: *Electrochemical Energy Storage for Renewable Sources and Grid Balancing*. Elsevier, 2015, pp. 253–268.

[83] Adam Z Weber et al. 'Redox flow batteries: a review'. In: *Journal of applied electrochemistry* 41.10 (2011), pp. 1137–1164.

[84] Hengyuan Hu et al. 'Development status, challenges, and perspectives of key components and systems of all-vanadium redox flow batteries'. In: *Future Batteries* 4 (2024), p. 100008.

[85] Abraham Alem Kebede et al. 'Techno-economic analysis of lithium-ion and lead-acid batteries in stationary energy storage application'. In: *Journal of Energy Storage* 40 (2021), p. 102748.

[86] Yushen Miao et al. 'Co-optimizing battery storage for energy arbitrage and frequency regulation in real-time markets using deep reinforcement learning'. In: *Energies* 14.24 (2021), p. 8365.

[87] Simone Coccato et al. 'A Review of Battery Energy Storage Optimization in the Built Environment'. In: *Batteries* 11.5 (2025), p. 179.

[88] Jinqiang Liu, Zhaoyu Wang and Chao Hu. 'Optimizing Size of Lithium-Ion Battery Combined with PV Generation'. In: *2019 IEEE Transportation Electrification Conference and Expo (ITEC)*. 2019, pp. 1–6. DOI: 10.1109/ITEC.2019.8790452.

[89] Tuhibur Rahman and Talal Alharbi. 'Exploring lithium-Ion battery degradation: A concise review of critical factors, impacts, data-driven degradation estimation techniques, and sustainable directions for energy storage systems'. In: *Batteries* 10.7 (2024), p. 220.

[90] Alberto Grimaldi et al. 'Techno-economic optimization of utility-scale battery storage integration with a wind farm for wholesale energy arbitrage considering wind curtailment and battery degradation'. In: *Journal of Energy Storage* 112 (2025), p. 115500.

[91] Andriy Vasylyev, Alberto Vannoni and Alessandro Sorce. 'Optimal dispatch of Li-Ion battery energy storage, reviewing and considering cycling and calendar ageing models'. In: *Applied Thermal Engineering* 265 (2025), p. 125597.

[92] K. T. Cho, M. C. Tucker and A. Z. Weber. 'Modeling and Optimization of Redox Flow Batteries'. In: *Energy Technology* 4.6 (2016), pp. 655–678. DOI: 10.1002/ente.201500422.

[93] Diana Cremoncini et al. 'Mixed Integer Linear Program model for optimized scheduling of a vanadium redox flow battery with variable efficiencies, capacity fade, and electrolyte maintenance'. In: *Journal of Energy Storage* 59 (2023), p. 106500.

[94] Xiong Binyu, Zhao Jiyun and Li Jinbin. 'Modeling of an all-vanadium redox flow battery and optimization of flow rates'. In: *2013 IEEE Power & Energy Society General Meeting*. 2013, pp. 1–5. DOI: 10.1109/PESMG.2013.6672599.

[95] Raden A. A. Ramadhan et al. 'Comparison of physical and machine learning models for estimating solar irradiance and photovoltaic power'. In: *Renewable Energy* 178 (2021), pp. 1006–1019.

[96] A. Dolara et al. 'PV hourly day-ahead power forecasting in a microgrid context'. In: *IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC)*. 2016, pp. 1–5.

[97] C. M. Omar et al. 'Day-ahead forecast for photovoltaic power using artificial neural networks ensembles'. In: *IEEE International Conference on Renewable Energy Research and Applications (ICRERA)*. 2016, pp. 1152–1157.

[98] J. Gaboitaiolelwe et al. 'Machine learning based solar photovoltaic power forecasting: a review and comparison'. In: *IEEE Access* 11 (2023), pp. 40820–40845.

[99] J. Polo et al. 'Exploring the PV Power Forecasting at Building Façades Using Gradient Boosting Methods'. In: *Energies* 16.3 (2023).

[100] G. Mitrentsis and H. Lens. 'An interpretable probabilistic model for short-term solar power forecasting using natural gradient boosting'. In: *Applied Energy* 309 (2022), p. 118473.

[101] A. Nespoli et al. 'Day-ahead photovoltaic forecasting: A comparison of the most effective techniques'. In: *Energies* 12.9 (2019), p. 1621.

[102] P. Li et al. 'A hybrid deep learning model for short-term PV power forecasting'. In: *Applied Energy* 259 (2020), p. 114216.

[103] F. Wang et al. 'A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework'. In: *Energy Conversion and Management* 212 (2020), p. 112766.

[104] Andreas Livera et al. 'Optimal development of location and technology independent machine learning photovoltaic performance predictive models'. In: *2019 IEEE 46th Photovoltaic Specialists Conference (PVSC)*. 2019, pp. 1270–1275.

[105] Julián Ascencio-Vásquez et al. 'Advanced PV performance modelling based on different levels of irradiance data accuracy'. In: *Energies* 13.9 (2020), p. 2166.

[106] Can Wan et al. 'Photovoltaic and solar power forecasting for smart grid energy management'. In: *CSEE Journal of Power and Energy Systems* 1.4 (2015), pp. 38–46. DOI: 10.17775/CSEEJPES.2015.00045.

[107] Z. M. Omer and H. Shareef. 'An adjustable machine learning gradient boosting-based controller for PV applications'. In: *Intelligent Systems with Applications* 19 (2023).

[108] A. Natekin and A. Knoll. 'Gradient boosting machines, a tutorial'. In: *Frontiers in Neurorobotics* 7 (2013), p. 21.

[109] U. Singh et al. 'A machine learning-based gradient boosting regression approach for wind power production forecasting: A step towards smart grid environments'. In: *Energies* 14.16 (2021), p. 5196.

[110] P. Sharma and J. Singh. 'Machine Learning Based Effort Estimation Using Standardization'. In: *International Conference on Computing, Power and Communication Technologies (GUCON)*. 2018, pp. 716–720.

[111] Minho Kim. 'Direct Short-Term Forecast of Photovoltaic Power through a Comparative Study between COMS and Himawari-8 Meteorological Satellite Images in a Deep Neural Network'. In: *Remote Sensing* 12.15 (2020), p. 2357. DOI: 10.3390/rs12152357.

[112] M. Pierro et al. 'Data-driven upscaling methods for regional photovoltaic power estimation and forecast using satellite and numerical weather prediction data'. In: *Solar Energy* 158 (2017), pp. 1026–1038. DOI: 10.1016/j.solener.2017.09.068.

[113] David P. Larson and Carlos F. M. Coimbra. 'Direct Power Output Forecasts from Remote Sensing Image Processing'. In: *Journal of Solar Energy Engineering* 140.2 (2018), p. 021011. DOI: 10.1115/1.4038970.

[114] Dazhi Yang and Richard Perez. 'Can we gauge forecasts using satellite-derived solar irradiance?' In: *Journal of Renewable and Sustainable Energy* 11.2 (2019), p. 023704. DOI: 10.1063/1.5087588.

[115] 'Addressing photovoltaic (PV) forecasting challenges: Satellite-driven data models for predicting actual PV generation using hybrid (LSTM-GRU) model'. In: *Energy Reports* 14 (2025), pp. 2141–2156. ISSN: 2352-4847. DOI: https://doi.org/10.1016/j.egyr.2025.08.034.

[116] Weatherbit. *Weatherbit*. https://www.weatherbit.io. Accessed: 2025-09-11.

[117] IRENA. *Renewable Energy Statistics 2020*. International Renewable Energy Agency, 2020. Chap. Renewable Hydropower (Including Mixed Plants).

[118] T. Polasek and M. Čadík. 'Predicting photovoltaic power production using high-uncertainty weather forecasts'. In: *Applied Energy* 339 (2023), p. 120989.

[119] Z. Hu et al. 'Improved multistep ahead photovoltaic power prediction model based on LSTM and self-attention with weather forecast data'. In: *Applied Energy* 359 (2024), p. 122709.

[120] D. Markovics and M. J. Mayer. 'Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction'. In: *Renewable and Sustainable Energy Reviews* 161 (2022), p. 112364.

[121] Costanza Luppi et al. 'Enhancing Photovoltaic Power Forecasting Using the LGB Model and Synthetic Features'. In: *IEEE Journal of Photovoltaics* (2025).

[122] J. C. Lam and D. H. W. Li. 'Correlation analysis of solar radiation and cloud cover'. In: *International Journal of Ambient Energy* 19.4 (1998), pp. 187–198.

[123] M. G. Nevins and J. N. Apell. 'Emerging investigator series: quantifying the impact of cloud cover on solar irradiance and environmental photodegradation'. In: *Environmental Science: Processes and Impacts* 23.12 (2021).

[124] F. Kasten and G. Czeplak. 'Solar and terrestrial radiation dependent on the amount and type of cloud'. In: *Solar Energy* 24.2 (1980), pp. 177–189.

[125] M. S. Ahamed et al. 'Cloud cover-based models for estimation of global solar radiation: A review and case study'. In: *International Journal of Green Energy* 19.2 (2022), pp. 175–189.

[126] R. Trinchero and F. Canavero. 'Machine learning regression techniques for the modeling of complex systems: An overview'. In: *IEEE Electromagnetic Compatibility Magazine* 10.4 (2021), pp. 71–79.

[127] D. Kumar et al. 'An ensemble model for short-term wind power forecasting using deep learning and gradient boosting algorithms'. In: *21st National Power Systems Conference (NPSC)*. IEEE, 2020, pp. 1–6.

[128] Y. Ledmaoui et al. 'Forecasting solar energy production: A comparative study of machine learning algorithms'. In: *Energy Reports* 10 (2023), pp. 1004–1012.

[129] N. Aksoy and I. Genc. 'Predictive models development using gradient boosting based methods for solar power plants'. In: *Journal of Computational Science* 67 (2023), p. 101958.

[130] Utpal Kumar Das et al. 'Forecasting of photovoltaic power generation and model optimization: A review'. In: *Renewable and Sustainable Energy Reviews* 81 (2018), pp. 912–928.

[131] Elnaz Yaghoubi et al. 'A Systematic Review and Meta-Analysis of Model Predictive Control in Microgrids: Moving Beyond Traditional Methods'. In: *Processes* 13.7 (2025), p. 2197.

[132] Ali Baniasadi et al. 'Optimal sizing design and operation of electrical and thermal energy storage systems in smart buildings'. In: *Journal of Energy Storage* 28 (2020), p. 101186.

[133] Navid Ghaffarzadeh et al. 'Optimal sizing of energy storage system in a micro grid using the mixed integer linear programming'. In: *International Journal of Renewable Energy Research-IJRER* 7.4 (2017), pp. 2004–2016.

[134] Michael Schimpe et al. 'Energy efficiency evaluation of a stationary lithium-ion battery container storage system via electro-thermal modeling and detailed component analysis'. In: *Applied energy* 210 (2018), pp. 211–229.

[135] Yuqing Chen et al. 'A review of lithium-ion battery safety concerns: The issues, strategies, and testing standards'. In: *Journal of Energy Chemistry* 59 (2021), pp. 83–99.

[136] Idris Temitope Adebanjo et al. 'A comprehensive review of lithium-ion battery components degradation and operational considerations: a safety perspective'. In: *Energy Advances* (2025).

[137] Yu Yang et al. 'Towards a safer lithium-ion batteries: A critical review on cause, characteristics, warning and disposal strategy for thermal runaway'. In: *Advances in Applied Energy* 11 (2023), p. 100146.

[138] Yan Wang et al. 'Advances in safety of lithium-ion batteries for energy storage: Hazard characteristics and active suppression techniques'. In: *Energy Reviews* 4.1 (2025), p. 100117.

[139] Judith A Jeevarajan et al. *Battery hazards for large energy storage systems*. 2022.

[140] Jens Conzen et al. 'Lithium ion battery energy storage systems (BESS) hazards'. In: *Journal of Loss Prevention in the Process Industries* 81 (2023), p. 104932.

[141] Aishwarya Parasuraman et al. 'Review of material research and development for vanadium redox flow battery applications'. In: *Electrochimica Acta* 101 (2013), pp. 27–40.

[142] Piergiorgio Alotto, Massimo Guarnieri and Federico Moro. 'Redox flow batteries for the storage of renewable energy: A review'. In: *Renewable and sustainable energy reviews* 29 (2014), pp. 325–335.

[143] Andrea Trovò et al. 'Redox flow batteries: A glance at safety and regulation issues'. In: *Electronics* 12.8 (2023), p. 1844.

[144] Vanadium Price. *China Connects World's Largest Redox Flow Battery System To Grid*. Consultato il 9 settembre 2025. Sept. 2022. URL: https://vanadiumprice.com/2022/09/30/china-connects-worlds-largest-redox-flow-battery-system-to-grid/.

[145] Abdullah M. Alharbi, Ibrahim Alsaidan and Wenzhong Gao. 'Sizing Li-ion BESS for Participation in Ancillary Services With Consideration of Degradation Cost'. In: *2022 IEEE Kansas Power and Energy Conference (KPEC)*. 2022, pp. 1–6. DOI: 10.1109/KPEC54747.2022.9814738.

[146] Xiaopeng Chen et al. 'An overview of lithium-ion batteries for electric vehicles'. In: *2012 10th International Power & Energy Conference (IPEC)*. IEEE. 2012, pp. 230–235.

[147] F.C. Walsh et al. 'The Characterization of Redox Flow Batteries: A Review'. In: *ChemPlusChem* 80.2 (2015), pp. 288–311. DOI: 10.1002/cplu.201402102.

[148] L. Thaller. *Electrically Rechargeable Redox Flow Cells*. Tech. rep. Report. NASA, 1974.

[149] NASA. *Redox Flow Cell Development and Demonstration Project, Calendar Year 1976*. Tech. rep. TM-73873. NASA Technical Memorandum. National Aeronautics and Space Administration, 1977.

[150] E. Sum and M. Skyllas-Kazacos. 'A study of the V(II)/V(III) redox couple for redox flow cell applications'. In: *Journal of Power Sources* 15 (1985), pp. 179–190. DOI: 10.1016/0378-7753(85)80024-2.

[151] Maria Skyllas-Kazacos et al. 'New All-Vanadium Redox Flow Cell'. In: *Journal of the Electrochemical Society* 133.5 (1986), pp. 1057–1058. DOI: 10.1149/1.2108706.

[152] M. Skyllas-Kazacos. 'Novel vanadium chloride/polyhalide redox flow battery'. In: *Journal of Power Sources* 124 (2003), pp. 299–302. DOI: 10.1016/S0378-7753(03)00668-2.

[153] Christina Roth, Jens Noack and Maria Skyllas-Kazacos, eds. *Flow Batteries: From Fundamentals to Applications*. Weinheim: Wiley-VCH, 2022. ISBN: 9783527347638.

[154] Vilayanur V Viswanathan et al. 'An overview of the design and optimized operation of vanadium redox flow batteries for durations in the range of 4–24 hours'. In: *Batteries* 9.4 (2023), p. 221.

[155] S. Roe, C. Menictas and M. Skyllas-Kazacos. 'A high energy density vanadium redox flow battery with 3 M vanadium electrolytes'. In: *Journal of The Electrochemical Society* 163.1 (2016), A5023–A5028. DOI: 10.1149/2.0071601jes.

[156] M. Krishna et al. 'Redox Flow Batteries: A Review on Technology, Applications and Trends'. In: *Journal of Energy Storage* 15 (2018), pp. 69–90. DOI: 10.1016/j.est.2017.11.005.

[157] Y. Zeng et al. 'Recent progress on the development of vanadium redox flow battery'. In: *Applied Energy* 255 (2019), p. 113756. DOI: 10.1016/j.apenergy.2019.113756.

[158] M. Risbud et al. 'Recent Advances in Vanadium Redox Flow Batteries'. In: *Batteries* 5.1 (2019), p. 24. DOI: 10.3390/batteries5010024.

[159] W Hill Balliet et al. 'Determining the profitability of energy storage over its life cycle using levelized cost of storage'. In: *Energy Economics* 142 (2025), p. 108174.

[160] ENTSO-E – European Network of Transmission System Operators for Electricity. *ENTSO-E: Transparency Platform and publications*. https://www.entsoe.eu/. Consultato il 4 settembre 2025. 2025.

[161] J. Noack et al. 'Techno-Economic Modeling and Analysis of Redox Flow Battery Systems'. In: *Energies* 9.8 (2016), p. 627. DOI: 10.3390/en9080627.

[162] Nicola Poli et al. 'Techno-economic assessment of future vanadium flow batteries based on real device/market parameters'. In: *Applied Energy* 362 (2024), p. 122954.

[163] Jianlin Li, Qian Wang and Jianhui Zhang. 'Design and development of large-scale vanadium redox flow batteries for engineering applications'. In: *Journal of Power Sources* 591 (2024), p. 233855.

[164] Qian Xu et al. 'Evaluation of Redox Flow Batteries Goes beyond Round-Trip Efficiency: A Technical Review'. In: *Journal of Energy Storage* 16 (2018), pp. 108–115. DOI: 10.1016/j.est.2018.01.005.

[165] William E. Hart et al. *Pyomo - Optimization Modeling in Python*. https://pyomo.readthedocs.io. Accessed: 2025-09-24. 2017.

[166] Free Software Foundation. *GNU Linear Programming Kit (GLPK)*. https://www.gnu.org/software/glpk/. Accessed: 2025-09-24. 2025.

[167] Hossein Lotfi and Amin Khodaei. 'Levelized cost of energy calculations for microgrids'. In: *2016 IEEE power and energy society general meeting (PESGM)*. IEEE. 2016, pp. 1–5.

[168] Tianmei Chen et al. 'Applications of lithium-ion batteries in grid-scale energy storage systems'. In: *Transactions of Tianjin University* 26.3 (2020), pp. 208–217.

[169] Wesley Cole and Akash Karmakar. *Cost Projections for Utility-Scale Battery Storage: 2023 Update*. Technical Report, National Renewable Energy Laboratory (NREL). 2023. DOI: 10.2172/1984976. URL: https://www.nrel.gov/docs/fy23osti/85332.pdf.

[170] Wiljan Vermeer, Gautham Ram Chandra Mouli and Pavol Bauer. 'A comprehensive review on the characteristics and modeling of lithium-ion battery aging'. In: *IEEE Transactions on Transportation Electrification* 8.2 (2021), pp. 2205–2232.

[171] Isaac Gwayi et al. 'A Review of Lithium-Ion Battery Empirical and Semi-Empirical Aging Models for Off-Grid Renewable Energy Systems Application'. In: *Engineering Reports* 7.5 (2025), e70169.

[172] Nils Collath et al. 'Aging aware operation of lithium-ion battery energy storage systems: A review'. In: *Journal of Energy Storage* 55 (2022), p. 105634.

[173] F. Lo Franco et al. 'Electric Vehicle Charging Hub Power Forecasting: A Statistical and Machine Learning Based Approach'. In: *Energies* 16.4 (2023), p. 2076.