Resilience Engineering Framework Integration in Off-Grid Renewable Energy Systems

Candidate
Pier Luca Anania

Supervisor
Prof. Andrea Micangeli

SAPIENZA
UNIVERSITÀ DI ROMA

Resilience Engineering Framework Integration for in Off-Grid Renewable Energy Systems

Faculty of Industrial and Civil Engineering
Department of Mechanical and Aerospace Engineering
Master's Degree in Mechanical Engineering

Pier Luca Anania
1653220

Supervisor
Prof. Andrea Micangeli

A.A. 2023-2024

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 Problem Contextualization

More than 1.4 billion people worldwide do not have access to electricity. Roughly 85% of these people live in rural areas and a large proportion live in Africa [1]. To date many utilities and governments have been unable to meet the energy needs of rural areas, as the focus has often been on meeting the demand of major industries or highly-populated urban areas such as the Nigeria’s area in Western Africa or the area around Lake Victoria on the Ugandan side. [2]

In the contemporary landscape of energy systems, minigrids have emerged as pivotal infrastructures, particularly in remote or off-grid areas, offering a decentralized and sustainable solution to electricity provision. However, ensuring the reliable operation of minigrids amidst diverse challenges poses a significant concern. Anomalies, ranging from equipment malfunctions to extreme weather events, can disrupt normal operations, leading to service interruptions and potential safety hazards. Addressing these challenges necessitates not only robust anomaly detection mechanisms but also a holistic approach that integrates principles of resilience engineering.

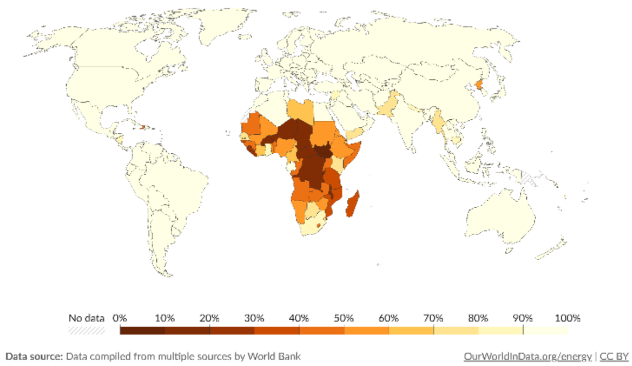

Fig.1: Share of the population with access to electricity (2020).
Data compiled from multiple sources by World Bank

An integrated approach utilizing the theoretical and practical principles of Resilience Engineering is crucial in a world of constant change, whether we are talking about phenomena relating to climate change, geopolitical instabilities or simply the reliability of a more or less complex energy system.

Being able to rely on continuous service is crucial in contexts of full electrification (think of the need to service critical infrastructure) as well as in contexts of rural electrification.

In a community where the energy supply is tied to a single source and its life and economy depend on it, it is more necessary than ever to define, from the earliest stages, a system capable of overcoming technical, operational and community shortcomings. [5]

1.2 Purpose and Objectives of the Thesis

The primary objective of this work is to develop a comprehensive understanding of how resilience engineering concepts can inform and improve anomaly detection strategies by leveraging insights from resilience engineering literature and methodologies, this study aims to enhance the robustness and adaptability of anomaly detection algorithms, thereby bolstering the overall resilience of minigrid operations.

Through the analysis of an Open-Source dataset concerning a photovoltaic production plant, an EDA Exploratory Data Analysis and the implementation of an Anomaly Detection algorithm will be carried out in order to highlight critical points in the system.

The aim of this thesis work is to structure a multidisciplinary and multiobjective approach in which the resilience engineering framework is applied to a photovoltaic energy production system. By fostering a deeper understanding of the interplay between resilience engineering and minigrid operations, this research endeavors to inform future strategies for enhancing the reliability and sustainability of decentralized energy systems.

1.3 Relevance of Resilience Engineering in Minigrids

Resilience engineering, a paradigm rooted in the fields of safety and systems engineering, emphasizes the ability of systems to adapt and recover from disruptions while maintaining essential functions. By shifting the focus from preventing failures to managing and mitigating their consequences, resilience engineering offers a promising framework for enhancing the performance and reliability of complex systems like minigrids.

The increasing demand for electricity and the need for sustainable energy sources have led to the development of various decentralized energy systems, including minigrids. However, these minigrids are often subject to disturbances and failures, which can have significant impacts on the communities they serve. Resilience engineering, which focuses on the ability of a system to adapt and recover from disturbances, is therefore highly relevant in the context of minigrids. This thesis will discuss the relevance of resilience engineering in minigrids, highlighting its importance in ensuring the sustainability and reliability of these energy systems.

Resilience engineering is a proactive approach to engineering that focuses on the ability of a system to anticipate, respond to, and recover from disturbances and failures [6]. It recognizes that disturbances are inevitable and that the goal is not to prevent them but to manage them in a way that minimizes their impact.

Minigrids, on the other hand, are small-scale, decentralized electricity distribution systems that serve a limited geographical area. They are often used in remote or rural areas where there is no access to the centralized grid. Minigrids can be powered by various energy sources, including fossil fuels, renewable energy, or a combination of both.

The relevance of resilience engineering in minigrids can be seen in several ways:

Improved System Reliability: Minigrids are often the sole source of electricity for the communities they serve. Any disruption in the supply of electricity can have significant impacts on the community's social and economic well-being. Resilience engineering can help improve the reliability of minigrids by ensuring that they can withstand and recover from disturbances quickly.

Cost-Effective: Resilience engineering focuses on managing disturbances rather than preventing them. This approach can be more cost-effective than trying to prevent all disturbances, which can be expensive and often not feasible. By managing disturbances effectively, minigrids can reduce the need for costly repairs and replacements.

Increased Sustainability: Resilience engineering can help increase the sustainability of minigrids by ensuring that they can adapt to changing conditions.

For example, minigrids that are designed with resilience engineering principles can better adapt to changes in energy demand, climate change, and technological advancements.

Improved Safety: Minigrids that are designed with resilience engineering principles can be safer for both the operators and the communities they serve. By anticipating and managing disturbances, minigrids can reduce the risk of accidents and injuries.

Resilience engineering is highly relevant in the context of minigrids. It can help improve the reliability, cost-effectiveness, sustainability, and safety of these energy systems. By focusing on the ability of minigrids to anticipate, respond to, and recover from disturbances, resilience engineering can ensure that minigrids can continue to provide essential electricity services to the communities they serve, even in the face of challenges and uncertainties. As the demand for decentralized energy systems continues to grow, the importance of resilience engineering in minigrids cannot be overstated.

1.4 Thesis Structure

The thesis work is organized with an initial review of the existing literature in Chapter 2 to build a solid and up-to-date background. It starts with an analysis of the Energy Access context, aligned with the Sustainable Development Goal (SDG) 7 Energy Access objective. This is followed by a study on the state of the art of minigrids in rural contexts. The components and characteristics useful for the discussion are defined. This is followed by a study of modern methodologies of Resilience Engineering, Exploratory Data Analysis and Anomaly Detection.

Chapter 3 describes the methodology by delving into the study context, the Resilience Engineering framework, and the integration of Anomaly Detection within it.

Chapter 4 addresses the Design and Implementation of the Anomaly Detection Algorithm, detailing the model and the code developed in the Python language.

Chapter 5 examines the Evaluation of Minigrid Resilience through the development of indicators and an impact analysis of the same.

Chapter 6 presents a final analysis and critical discussion of the results obtained, highlighting limitations and potential developments.

The work concludes with acknowledgments in Chapter 8 and the Bibliography in Chapter 9.

2. Literature Review

2.1 Energy Access Context

The Energy Access Context delves into the intricate landscape surrounding energy accessibility. This section examines the multifaceted factors influencing the availability and affordability of energy resources worldwide. From socio-economic disparities to infrastructural challenges, understanding the context of energy access is paramount in addressing global energy inequality.

2.1.1 Energy Planning

There is no universal definition of the term "Energy Access." IEA (2011) gives the following definition: "a household having reliable and affordable access to clean cooking facilities, a first connection to electricity and then an increasing level of electricity consumption over time to reach the regional average." However, the definition implicitly assumes the regional average level of consumption as the acceptable minimum need which can be problematic due to its potential for encouraging wasteful consumption and perpetuation of unsustainable lifestyles. [3]

Globally, approximately 759 million individuals, constituting 1 out of every 10 people, lack access to essential electricity for illuminating their homes, preserving perishable food items, or mitigating the effects of escalating temperatures. Roughly 2.6 billion individuals are compelled to resort to polluting biomass sources like charcoal, coal, and animal waste for cooking purposes. These statistics present an intolerable reality.

In Sub-Saharan Africa and Asia, the largest disparities in electricity and clean cooking accessibility are observed across 20 countries. These regions also contribute to the 80 percent of nations worldwide that grapple with inadequate electricity provision. The absence of access to clean, modern energy undermines efforts to achieve Sustainable Development Goals (SDGs) aimed at poverty alleviation, educational enhancement, and public health amelioration. For instance, replacing antiquated stoves and open fires could prevent the deaths of 800,000 children annually, who succumb to indoor air pollution exposure. Hence, the imperative of SDG7 is to address these energy disparities by 2030. [4]

Focusing on the African continent in pursuit of the ambitious goal of achieving universal access to modern energy services across Africa by 2030, it becomes imperative to explore diverse pathways within the electricity sector. Such exploration not only aids policy-makers and investors in making informed decisions but also plays a pivotal role in shaping the design of power systems. [7]

Holistic energy systems planning endeavors to ensure that energy-related policy and investment choices encompass all viable options on both the supply and demand sides, aligning with broader national objectives such as sustainable development. However, a fundamental prerequisite is the establishment of robust national energy planning capability.

Energy planning capacity serves as a cornerstone, enhancing a country's capacity to anticipate and adapt to rapid changes while capitalizing on emerging opportunities and addressing new challenges. This asset appreciates over time as experts accumulate practical experience, enrich the local knowledge repository, and foster collaborations with stakeholders across various sectors. Historically, inadequate national planning capacity has resulted in suboptimal policy and investment decisions, contributing to unequal access to modern energy services.

Furthermore, energy planning transcends national boundaries, particularly for smaller nations with limited energy resource potentials, such as hydropower. Collaborative ventures involving infrastructure-sharing with neighboring countries offer the potential for economies of scale, highlighting the interconnected nature of energy planning across geopolitical borders.

Over the past two decades, numerous developing nations have embraced extensive policies advocating for liberalization and privatization, often under the influence of major international funders and development organizations. While these policies have occasionally bolstered the operational efficiency of individual national utilities, their impact on expanding energy access has been modest at best. This is primarily due to the fact that catering to the electricity needs of the most marginalized populations isn't financially lucrative for utilities.

This discourse mirrors the ongoing dialogue within OECD countries over the same period, where the outcomes have been similarly varied. The purported advantages of liberalizing these predominantly fragile markets remain ambiguous. In instances where liberalization has been ideologically imposed on these nations, it often proves detrimental, despite originating from well-meaning intentions.

Similar to many sectors of public policy, energy policy formulation heavily relies on analytical models. However, these models exhibit significant variations in outputs, temporal and spatial scopes, sophistication levels, terminology, underlying assumptions, system boundaries, and theoretical frameworks. Consequently, the findings generated from these analyses necessitate substantial filtration and translation to effectively inform the design and implementation of governmental policies.

In line with this perspective a noticeable disparity between the inquiries posed by policymakers and the outcomes derives from modeling exercises. Within the realm of energy policy, power system analyses represent a subset of broader energy system modeling endeavors. Integrated resource planning models (IRP) commonly serve as pivotal tools within the power sector for strategic decision-making processes.

Power system analyses, management, and planning encompass various timeframes, spanning from sub-second activities like load balancing to multi-decade projections for capacity expansion. Fundamental to this planning is a set of electricity demand projections, forming the basis for capacity expansion strategies. Such planning often centers on least-cost optimization methodologies, considering a spectrum of constraints such as existing infrastructure conditions, financial accessibility, environmental policies, and energy security imperatives.

Across governmental planning agencies and utilities globally, an array of modern mathematical techniques is commonly employed. These range from fuzzy logic and evolutionary programming to mixed-integer linear programming and multi-objective optimization. Recent research in this domain reveals a discernible trend towards incorporating uncertainties and adapting to liberalized market dynamics.

However, for many power systems in sub-Saharan Africa, excessively sophisticated methodologies may not be imperative initially for initiating generation and infrastructure planning processes.

Energy demand projections constitute a pivotal element in the majority of planning initiatives. Various tools and methodologies of differing degrees of complexity, as outlined in Table 1, are employed to forecast future demand.

Type	Description
Trend method	Non-causal model, i.e. it does not explicitly explain how the projected variable is determined, which is purely a function of time (e.g., x% increase per year).
End-use method (or engineering based method)	Approach based on energy usage patterns of appliances and systems.
Agent-based models	Class of computational models for simulating the actions and interactions of autonomous agents (both individual and collective entities) with a view to assessing their effects on the system as a whole. The models simulate the simultaneous operations and interactions of multiple agents, in an attempt to re-create and predict the appearance of complex phenomena.
Time series method	Projections solely based on historical patterns in the data.
Econometric method	Standard statistical tools are employed to produce a mathematical representation of the energy demand as a function of a series of variables (e.g., population, GDP). The functions derived can then be used to project the demand into the future, assuming that the causal relationships remain unchanged over time. Alternatively causal relationships are guided by normative of policy objectives.
Neural network techniques	Techniques which are able to capture and represent complex input/output relationships, both linear and non-linear. The advantage is the ability to learn these relationships directly from the data being modeled. Usually used for short-term load forecasting.

Table 1: Selected methods for energy demand forecasting (adapted from: McDowall & Eames (2006) and Thomas (2006)) [7]

Each approach to energy demand projection possesses distinct strengths and weaknesses. The selection of the appropriate method hinges on several factors, notably the nature and availability of underlying data, as well as the purpose of the analysis and the timeframe involved.

In many long-term planning endeavors conducted in sub-Saharan Africa (SSA), demand projections often rely on econometric relationships tied to income (GDP) and population growth projections, coupled with elasticity relationships. Furthermore, certain methodologies incorporate explicit terms for household connections and large point demands. For instance, consider Equation 1 as depicted in PIDA (2011):

$$D_t = D_{t-1}\left(\epsilon \frac{\Delta GDP}{GDP} + 1\right) + kC_t + \Delta M_t \tag{1}$$

Where:

- D is the unconstrained demand
- ϵ is the GDP elasticity of electricity demand
- k is the average annual consumption of electricity of one household
- C_t is the number of new connections in a year t
- M_t is the additional demand from new large demand points

In contexts where a significant portion of the population lacks access to electricity services, traditional techniques reliant on aggregates like GDP and exogenous inputs such as future annual grid connections of households may not be optimally suited. In such scenarios, alternative approaches, such as solving for a future goal and back-casting, rather than forecasting based solely on historical trends, become necessary. Ensuring that the analytical approach aligns with the specific policy and investment inquiries at hand is paramount. It has been contended that in severely supply-constrained electricity systems, demand projections hold less significance compared to capacity expansion planning and associated financing. Put differently, in typical developing country settings, additional supply tends to stimulate its own demand.

2.1.2 Historical Energy Trends

Sub-Saharan Africa grapples profoundly with a dire lack of access to electricity and subpar quality of supply, characterized by issues of cost and reliability where infrastructure exists. An estimated 580 million individuals across the continent lack access to electricity, with the majority residing in rural areas (IEA, UNDP, and UNIDO, 2010). The electrification rate in SSA stands at approximately 30%, with urban areas showing higher rates at 60% compared to rural areas at 14% (IEA, UNDP, and UNIDO, 2010).

Numerous sources offer comprehensive analyses of the energy landscape in Africa, such as Eberhard et al. (2011) [8]. Recent academic literature on Africa's power systems prominently features discussions concerning solar power in North Africa. Additionally, much of the literature pertaining to the power sector in SSA understandably focuses on the Republic of South Africa (RSA). Nonetheless, there exists a dedicated cohort of researchers who focus on SSA as a whole or on specific countries within the region. Despite this, the literature on power sector scenarios in sub-Saharan Africa remains relatively sparse.

The total average per capita consumption in SSA (excluding RSA) is around 155 kWh (based on EIA data). These figures are minute compared to Sth. Africa where this value is approximately 4770 kWh per capita or other OECD countries.

The installed capacity in Africa will need to grow by more then 10% just to meet Africa’s suppressed demand, keep pace with projected economic growth and provide additional capacity to support efforts to expand electrification. Most new capacity would be used to meet non-residential demands from the commercial and industrial sectors. [7]

Figure 2 shows the total electricity generation capacity installed per million persons (MW/mln) in several regions. It is argued that is a relatively rough metric as it does not take into account a number of different and crucial parameters, including: T&D Transmission and Distribution losses, load patterns, locational constraints, intermittency, temporal reserve, availability, operating efficiency, and outage rates. Compared to the other world regions, the ratio of electricity generation capacity per million inhabitants is low in Africa, particularly in sub-Saharan Africa. The figure for SSA (excluding RSA) was roughly 129 MW/mln in 2008 only considering people with electricity access; if the entire population is included, the total is about 40 MW/mln.

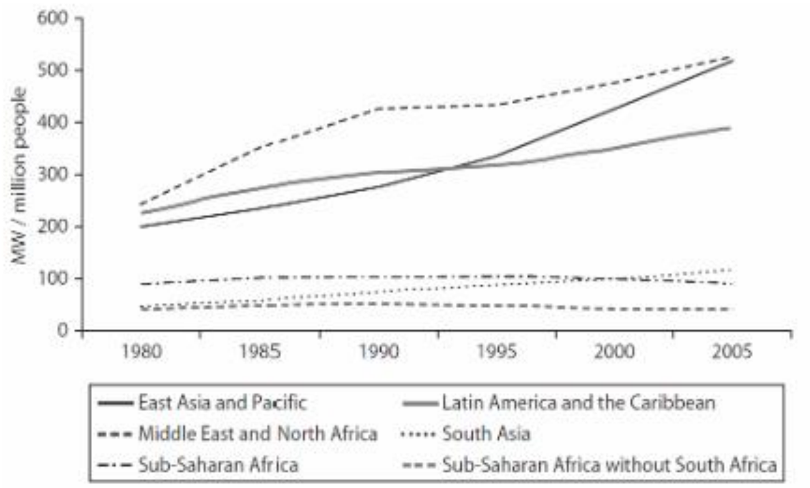

Fig.2: MWs installed per one million by region [8]

History presents compelling evidence that significant increases in the percentage of households with access to electricity can be achieved over relatively short periods. For instance, electrification rates surged notably in several countries, including the USA and UK during the early 20th century, and more recently in China, Brazil, and Thailand (refer to Figure 3).

As an illustrative case, Thailand witnessed a remarkable transformation, with the percentage of the population with access to electricity escalating from approximately 25% to nearly 100% within a decade. However, for most nations, this transition typically spans at least three decades, if not longer. Across these countries, prioritizing electrification, particularly in rural areas, stemmed from high national priorities driven by economic development or equity objectives.

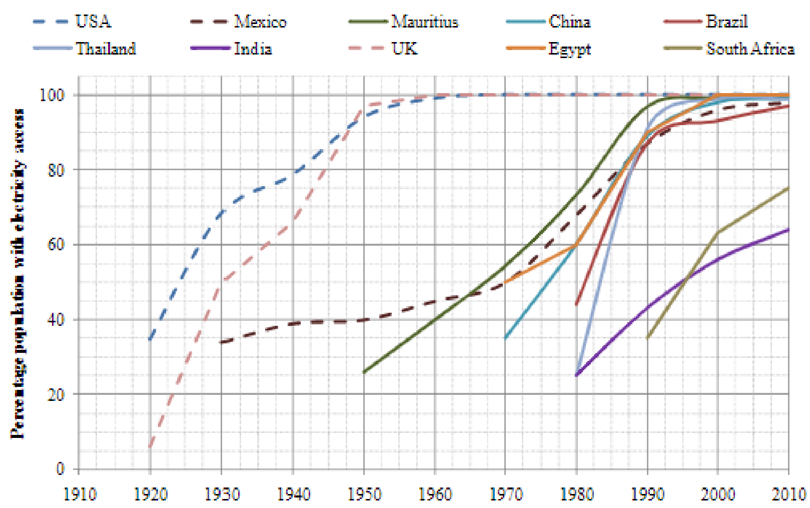

Fig. 3: Evolution of household electrification over time in selected countries [7]

While several countries in sub-Saharan Africa (SSA) have experienced remarkable growth, roughly quadrupling their installed capacity over the past two decades, these advancements mostly originated from a relatively modest initial installed capacity. However, the majority of countries in the region have witnessed sluggish growth or even a decline in installed capacity.

On average, installed electricity capacity in SSA (excluding RSA) has expanded relatively steadily at a rate of around 1.7% per annum. Examining the historical growth (or contraction) rates in African countries (refer to Figure 4) yields valuable insights for several reasons. First, it elucidates that there is no discernible pattern indicating an overall increase in growth rates over time. Despite the growing recognition of the pivotal importance of energy, particularly electricity, efforts to augment generating capacity have not shown a consistent upward trajectory in recent years.

Nevertheless, there are early indications suggesting a potential acceleration in the expansion of Africa's generation capacity. Data on donor commitments to power projects suggest that, over the last five years, an average annual commitment of 3 GW of generation projects has been made. Additionally, the Annual Report of the Infrastructure Consortium for Africa 2010 highlights a significant increase in member commitments to energy projects in sub-Saharan Africa, rising from USD 1.2 billion in 2006 to USD 8.0 billion in 2010.

Secondly, while the growth rate exhibits a wide range of values, it typically falls between 0% and 10%, with the bulk of countries experiencing growth rates between 0% and 5%. Thirdly, the variability of the change in installed capacity is high, although it has been decreasing over time, especially in recent years. Finally, the graphical representation indicates that countries with larger systems (depicted as red dots in the figure), characterized by greater existing capacity and transmission and distribution grids, tend to expand their capacity more rapidly than countries with medium and small electricity systems. In fact, with a few exceptions, countries with smaller electricity systems (represented by blue dots in the graph) exhibit relatively low growth rates or even negative growth, particularly towards the end of the 1990s.

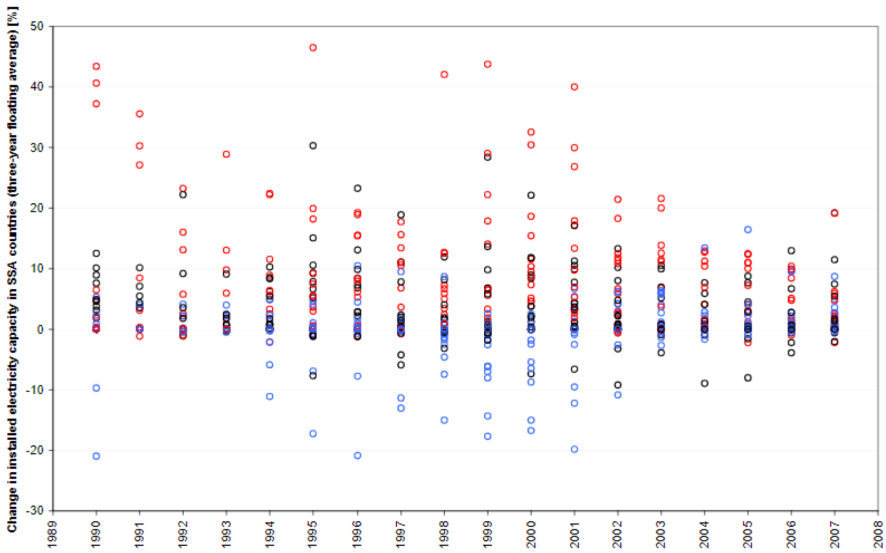

Fig. 4: Rate of increase (or decrease) in installed electricity capacity (with three year floating average) in SSA countries arranged by tertile (red, black and blue dots features countries with relatively large, medium, and small generating capacity, respectively, in 2008). Data: authors' compilation from EIA

Of course, SSA countries and regions are well aware of the problems of energy access, both in terms of quantity and quality, and have developed national targets and regional plans. UNDP and WHO (2009) calculated that 68 developing countries have electricity targets.

2.1.3 Prospects for Africa

In this section is briefly considered some of the datasets and projections for the power sector in Africa. For an initial sense of scale, using EIA data, Africa has a current installed generating capacity of about 122 GW, SSA (excluding RSA) had 31 GW. This compares roughly to 28 GW in Argentina.

Africa is included in the major energy outlooks from the International energy Agency (IEA), the US dept of Energy's Energy Information Agency (EIA), British Petroleum (BP) and other international committee. Each dataset has different levels of descriptive information coverage and aggregation. We primarily relied on the EIA dataset as it was the most transparent and complete in terms of accessible country time-series data. It is useful to look at results of these high level global exercises to get a sense of the numbers being fed into the *Global Energy Dialogue*.

Most of the African sub-regions have carried out forecasting exercises for peak energy demand, commonly both in terms of peak demand (or generation capacity) and consumption (or generation). Those projections are normally based on studies conducted at the national level.

Despite forecasting methods that vary considerably, the regional plans and related documents entail a wealth of quantitative information that is all too often underutilized in further analysis and planning.

The New Partnership for Africa's Development (NEPAD), the Southern African Development Community (SADC), the Forum of Energy Ministers in Africa (FEMA), the Economic Community of West African States (ECOWAS), the East African Community (EAC) and the Central African Economic and Monetary Community Commission (CEMAC), among others, have produced strategies for electrification and increasing access to modern fuels.

A closer look at some of the regional forecasts in the interests of comparison is useful. A SAPP electricity demand forecast to 2025 shows a projected annual growth of about 2% (SAPP, 2010); the annual growth rates are projected to be higher outside RSA. Nexant shows projected WAPP average growth of 7.6% (ranging from 5-12.6%). The EAC/EAPP Demand Forecasts show very large ranges in forecasted annual growth. They provide very detailed analysis of each country's national forecasts and the extend them to 2038 where appropriate. Interestingly, the forecasts for many of the countries show the same kind of exponential growth shown in next figure and reflect more typical trend or regression-based forecasts for "low and base" cases. Figure 5 shows the forecast to 2038 (in MW) for peak demand in Kenya, including showing sharp growth in the "High Case" from 1 GW to over 18 GW to 2038.

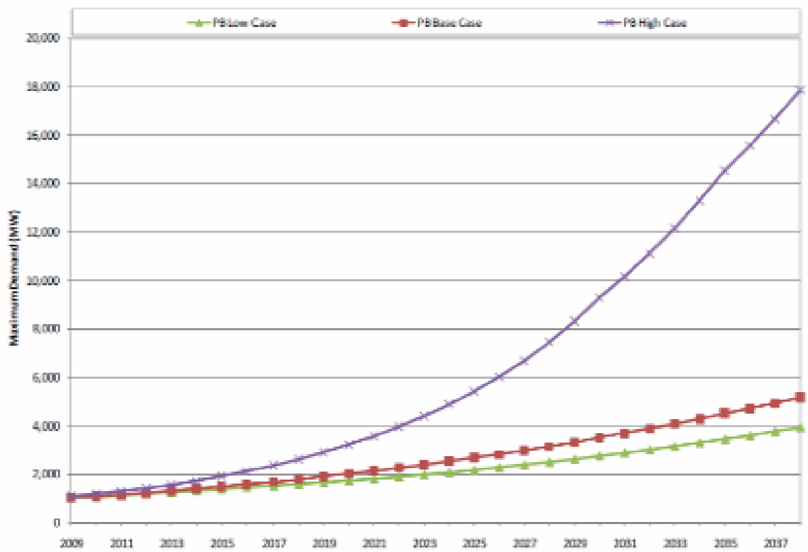

Fig.5: Peak demand forecasts for Kenya [7]

In [8] are shown several scenarios for Africa. They considered three type of demand: market, suppressed and social to help create three scenarios (constant access, regional target and national targets). The overall average annual electricity demand growth rate was estimated at 5.8%.

The objective of the context is to establish an infrastructure development programme articulated around priorities and phases and, prepare an implementation strategy and process including, in particular a priority action plan. The peak demand projections from initial Programme for Infrastructure Development in Africa (PIDA) shows an average 6.7% growth (with regional annual growth rates ranging from about 6-9%) over the period 2009-2040. The initial results assume that the access rate will increase from 42% in 2009 to 65% in 2030; these rates are projected to be similar in 2040.

The African Development Bank undertook a universal access scenario assessment through 2030. In Table 2 is shown the results of the capacity additions estimated. Without South Africa the total equals 102 GW, so approximately an average of 6% annual growth.

	Generating Capacity [GW]		
	Net	***Replacement***	***Total***
Northern Africa: 5 Nations	60	22	82
South Africa	47	30	77
Sub-Saharan Africa: 41 Nations	82.5	19	**102**
Island states: 6 Nations	2.5	1.5	4
Africa	192	72	265

Table 2: Universal Energy Access scenario to 2030 (African Development Bank)

2.1.4 Generation Technology

Now we investigate the various projections in terms of technology and energy resources. A special attention is given to renewable energy potentials, following the sustainable energy goal propose by the United Nations, in order to give a sense of scale to the possibilities.

In [8] is reported that over 900 TWh (approximately 220 GW installed capacity) of economically viable hydropower potential in Africa remains unexploited, located primarly in the Democratic Republic of Congo, Ethiopia, Cameroon, Angola, Madagascar, Gabon, Mozambique and Nigeria.

Similarly, the Intergovernmental Panel on Climate Change (IPCC) estimates the technical hydropower potential at 1174 TWh (or 283 GW of installed capacity), only eight percent of which has been developed. Interestingly, this unused potential is about ten times the current installed generating capacity in SSA if RSA is excluded.

The International Renewable Energy Agency (IRENA) is now designing future renewable energy scenarios. The focus of their work will be on providing detailed, regional specific technology information with a clear focus on renewable energy. The following Table shows that the technical potential for renewables is enormous, and largely untapped in Africa. The accounting of biomass remains contentious; still, even using conservative assumptions, the potentials are significant.

Region	Wind [TWh/yr]	Solar [TWh/yr]	Biomass [EJ/yr]	Geothermal [TWh/yr]	Hydro [TWh/yr]
East	2 000 – 3 000	30 000	20 – 74	1 – 16	578
Central	-	-	49 – 86	-	1 057
North	3 000 – 4 000	50 000 – 60 000	8 – 15	-	78
South	16	25 000 – 30 000	3 – 101	-	26
West	0 – 7	50 000	2 – 96	-	105
Total Africa	5 000 – 7 000	155 000 – 110 000	82 – 372	1 – 16	1 844

Table 3: Technical potential for renewable energy in Africa by region (IRENA)

In Figure 6 is used a ternary graph to plot selected (international organization) projections in terms of electricity production in Africa by types of energy sources, namely coal and oil, renewables, and low-carbon (nuclear and gas). Such representation allows visualizing the foreseen transition in the electricity generation and corresponding technological and resources shift. The portfolio of generation types critically impacts power system design and operation (including the amount of total installed capacity required because of issues such as intermittency, ramping rates, and inertial response). All of the projections foresee a decrease, in relative terms, of carbon intensive resources in Africa in the coming two decades, including those scenarios without an explicit focus on climate change mitigation.

Also, most projections feature an increase in low-carbon technologies in a first phase, before the share of renewable picks up significantly.

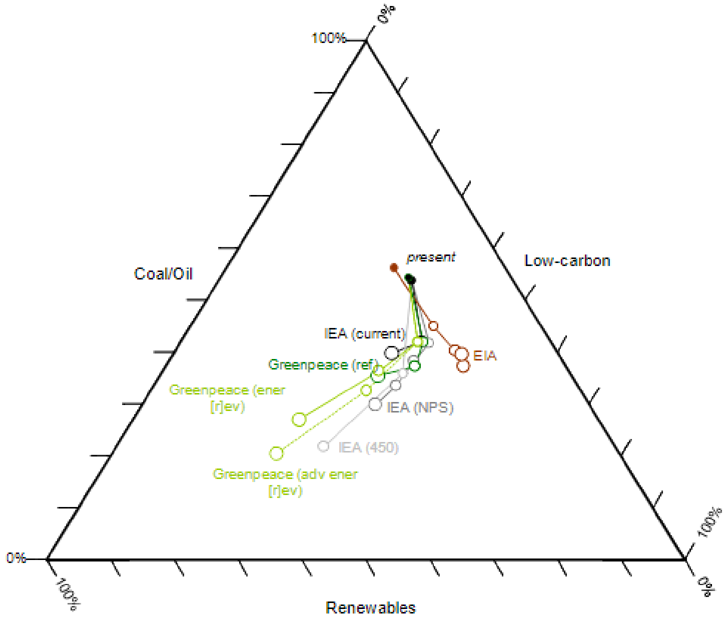

Fig.6: Projections of electricity generation in Africa by types of by different organisations, 2010-2030. Note: the size of the dots is proportional to the total electricity generation projected; with present estimates (filled dots), estimates in 2030 (last dot of each scenario), and intermediary estimates. Data: own compilation from IEA WEO 2010, EIA IEO 2010, and Greenpeace 2010

2.1.5 Scenarios to 2030

Using simple heuristics, it has been calculated "back of the envelope" electricity generation capacity requires in SSA (excluding RSA) to 2030 under various electricity access level assumptions (see table 4). It is important to note that these scenarios are not limited to household demand, but for the entire economy. In the first two scenarios it is separated the number of people without access (electricity poor) from those with access (non-electricity poor), and each category arrives at a different level of access in 2030. In the two other scenarios the entire 2030 population is brought to a single average level of access. Of course, such results are highly stylized and would, in themselves, not properly consider issue such as: intermittency, system operation, ramping etc..

The results of this evaluation are astonishing in term of the required growth rates and installed capacity. As an example, just to reach the Moderate Access case where the population has between 200-400 MW/mln requires a total of around 374 GW of installed capacity- about twelve times current levels. This implies around a 13% annual growth rate for the next 20 years as compared to 1.7% for the past 20 years. The other scenarios show that bringing access to the projected SSA (excluding RSA) population in 2030 would take approximately 500 GW to reach an average of 400 MW/mln (Full Access) and to reach 800 MW/mln (Full Enhanced Access) would double this requirement.

The result assumes much higher level of access than much of the literature that focuses solely on "basic needs" at the household level.

Level of access	2010 [GW]	2030 [GW]	Implied average annual growth rate 2010-2030	Scenario Name
Population - electricity poor, million	573	638	0.5%	
Population - non electricity-poor, million	240	615	4.8%	
electricity poor: 0 MW/mln non electricity-poor: 129 MW/mln[38]	31	79	4.8%	*Business As Usual*
electricity poor: 200 MW/mln non electricity-poor: 400 MW/mln		374	13.3%	*Moderate Access*
full population: 400 MW/mln		501	14.9%	*Full Access*
full population: 800 MW/mln		1,002	19.0%	*Full Enhanced Access*

Table 4: Estimates for installed electricity generation capacity required (in GW) in SSA (excluding RSA) under various access level (MW/mln) assumptions.

The next figure provides a simplified overview of several scenarios as well as projections. In addition to plotting the *Moderating Access* and *Full Access* scenarios from Table 4, it includes: a *50% Access* scenario that assumes that 50% of the population will have access at a rate of 400 MW/mln, along with two statistically derived projections based on historical data. *GPD regression* represents a regression analysis using GDP as the independent variable (with double exponential smoothing of historic data) and results in about 70 GW in 2030. The *Trendline* is a historically-based extrapolation, and projects about 43 GW in 2030.

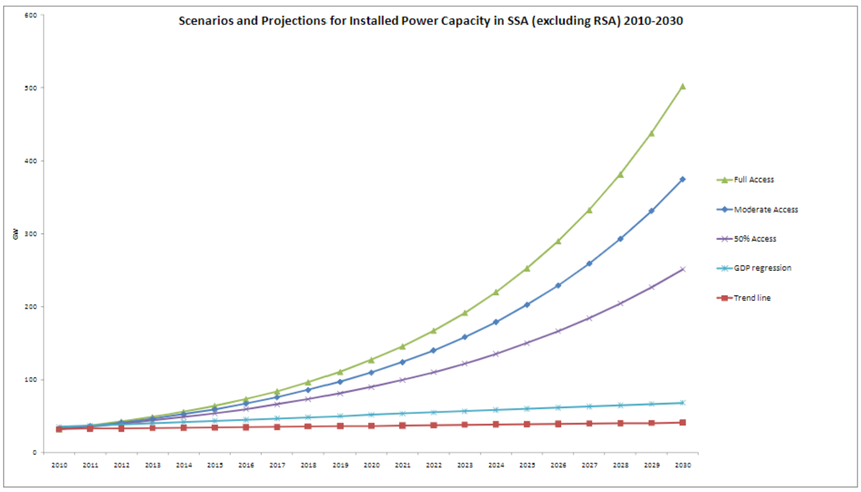

Fig. 7: Scenarios and projections of installed capacity to 2030 for SSA (excluding RSA)

It is also useful to consider how to “jump-start” from historic trends to, as an example, the *Low Access* case. A few well-designed large projects allow very high initial growth levels to help give confidence to the sector for an extended period of growth. For instance, the proposed Grand Inga hydroelectric project (in the Democratic Republic of Congo) could reach almost 40 GW in scale. Inga then would, theoretically, provide a significant short-term contribution to the additional capacity required. Likewise, some Nigerian projection show very high levels of short-term growth in generating plants. A few such large-scale projects might also provide the necessary impetus for transmission projects. High levels of growth in smaller or distributed generation projects would also likely support the necessary momentum.

Finally, while it is variable to illustrate what it would mean to meet a target of 100% electrification by 2030, it is also important to acknowledge that this target seems ambitious. As noted above, 30-40 years is likely a more realistic range based on the historical evidence presented, particularly given the following considerations:

- The final segment from 90%-100% access is necessarily slower due to increasing marginal costs and technical difficulties
- In addition, for Africa to meet the universal electrification target 47 countries would need to do simultaneously.

Building on *Full Access* scenario, it has been briefly examined the goal “spread” evenly across the sub-regional power pool level (the Eastern Africa Power Pool (EAPP), Southern Africa Power Pool (SAPP), Western Africa Power Pool (WAPP), and the Central African Power Pool (CAPP)).

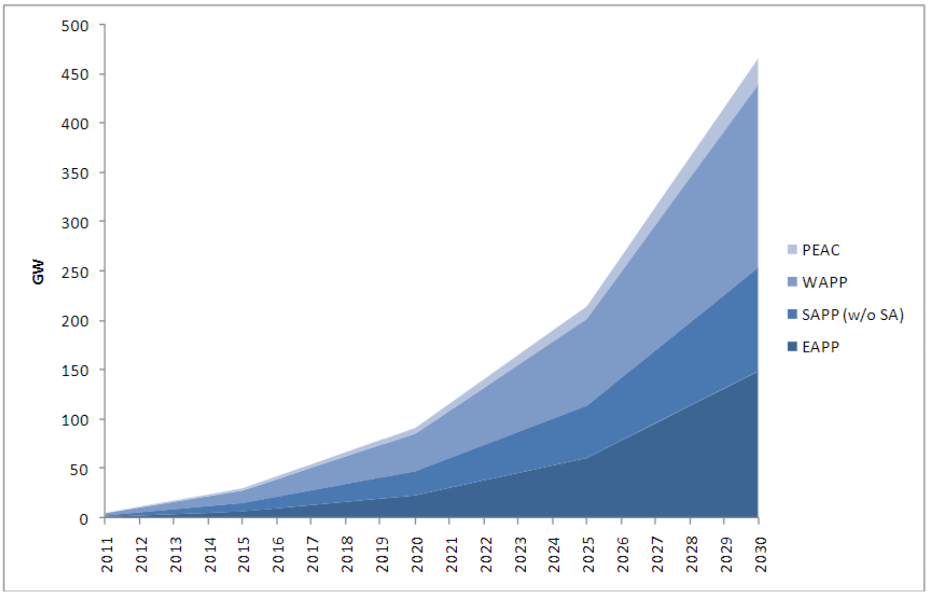

Fig. 8: Additional capacity needed to reach 400 MW/mln by region

Initial results show that the WAPP has the largest total capacity in additions at 186 GW, the EAPP has 149 GW, the SAPP (excluding RSA) has 105 GW, and the CAPP has 27 GW. On an average annual basis then, SSA (excluding RSA) must add about 23 GW per year in additional capacity (EAPP: 7.4 GW, SAPP (excluding RSA): 5.2 GW, WAPP: 9.3 GW and CAPP: 1.4 GW) – equivalent to a little more than a Three Gorges Dam (22.5 GW) sized project each and every year through 2030.

2.2 Minigrids Design Overview

Our modern society is highly dependent on the electrical grid and major outages have severe consequences. A reliable source of power is especially important for campuses (including college campuses, business parks, etc.), military bases, and other areas with critical municipal functions (such as hospitals, police, and fire stations), where public safety may be compromised by a lack of electrical power. Although backup generation is common at critical facilities, failure of backup generation resources is quite common due to lack of maintenance or insufficient fuel supplies. Advanced microgrids can be an effective solution for power delivery to critical infrastructure.

We consider a "microgrid" as an integrated energy system consisting of loads and generation operating as a coherent unit. Microgrids may operate either in parallel with, or islanded form the main electric grid, and may switch between these two states. A simple microgrid might involve minimal design effort and employ a simple design, such as only a critical load paired with a backup generator.

Simple designs are typically inefficient solutions when considering all critical loads and possible threats to a given system. An "advanced microgrid" is one that is designed using Sandia National Laboratories' Energy Surety Design Methodology (ESDM), which is a systematic process to maintain or enhance the attributes of: safety, security, reliability, sustainability, cost effectiveness, and resilience. Key components of advanced microgrid design include identifying and prioritizing critical assets, defining design basis threats, and establishing performance goals. [9]

Maintaining local power delivery during extended main electric grid outages has become increasingly important as more customers and services rely on electric power. This is highlighted in Figure 9, which shows that several of the worst blackouts in the world in terms of customer hours lost have occurred in the last 20 years.

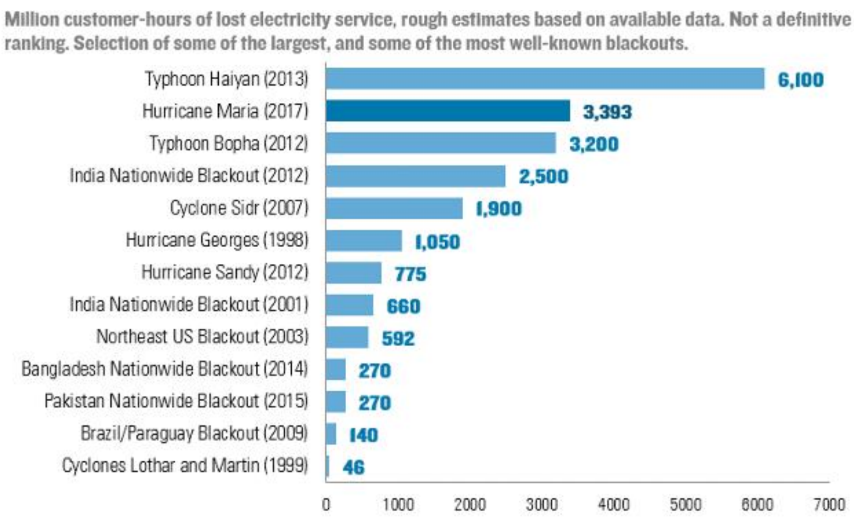

Fig. 9: Million customer-hours of lost electricity service, rough estimates based on available data. Selection of some of the largest and well-known blackouts. Source: DOE, National Academies, news reports, government statistics, academic, literature and Rhodium estimates

Due to interdependencies, extended power outages have cascading impacts on productivity, safety, and public health. Loss of power to a water treatment plant for an extended period will deplete reserves, impacting not only public health, but also firefighting and water for industrial uses. Outages to communications infrastructure due to lost power impacts the ability to dispatch emergency services, to coordinate mitigation efforts such as clearing debris, and to communicate with customers. Traffic signal outages and an inability to pump fuel due to power outages can cripple transportation.

These issues highlight how important it is for communities to consider options such as advanced microgrids to improve the design, operation, and management of their energy system infrastructure to minimize the impacts of extended electric grid outages.

The need is for energy surety: energy systems that are safe, secure, reliable, and designed in a way that provides energy system operational assurance during routine and extended impact events caused by accidents, natural disasters, or intentional attacks. [5]

2.2.1 Main Electric Grid

Most electric customers are served by a main electric grid. Main electric grids may span entire continents or may cover only a small island. These electric grids typically consist of the four components shown in Figure 10: generation, transmission, distribution, and customers, although smaller systems may not have significant transmission components.

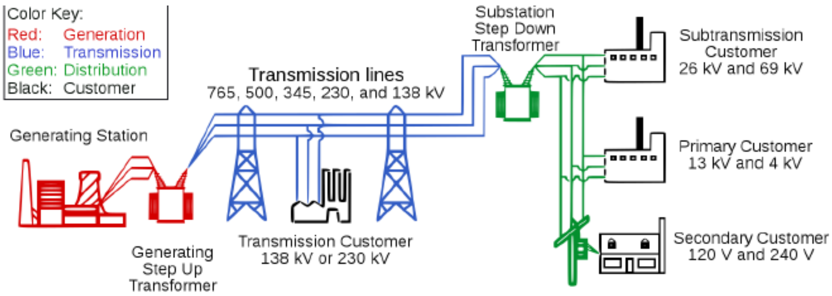

Fig. 10: Basic components of an electric grid (Image from FERC report: https://www.ferc.gov/industries/electric/indus-act/reliability/blackout/ch1-3.pdf.)

Common types of generation include coal (~27% of worldwide generation); natural gas (~27%); nuclear (~18%); hydroelectric (~13%); wind, solar, and geothermal (~10%); biofuel (~3%); and oil (~2%) power plants (IEA 2017 provisional electricity production by source: https://www.iea.org/statistics/electricity/). Electric grids have typically been operated with large generating stations, with power generally flowing from generating station to customer load, as illustrated in Figure 10. However, the growth of renewable energy such as wind and solar is increasingly spreading out the generation. Utility-scale wind and solar plants may be connected to transmission lines at myriad locations across the electric grid, and residential and commercial solar often exist on the distribution system, with residential systems often behind the customer meter. Generation may be owned by an electric utility or may be owned by a private entity that contracts with the utility, such as through a power purchase agreement.

Transmission systems are networks of transmission lines designed to transport energy over long distances with minimal power losses. They are often complex mesh networks with multiple redundant paths which can be utilized in the event of a single node failure.

Transmission lines are typically administered by a regional transmission organization or an intendent system operator. Careful attention is paid to balancing load and generation, maintaining a set frequency, and balancing the voltage between the three different phases.

Distribution systems complete the delivery of power to customers. The backbone of distributions systems is a high voltage "primary" system which, similar to transmission but at a lower voltage, transports the power closer to the customer. At or near the customers, distribution transformers reduce the voltage to customer-appropriate levels (such as 240V/120V). This lower voltage system is called the distribution "secondary" system and connects the low-voltage side of the distribution transformer to the customer meter. Electric utilities manage distribution systems, ensuring that power is delivered to customers at safe voltage levels.

Customers can range from large industrial complexes to single-family homes. Customer voltages will vary depending on the size of the load and types of equipment used by the customer.

Distribution system equipment including transformers and wires leading to the customer must be sized appropriately for the loads. A special case of customer is one that has generation behind the meter, such as a rooftop photovoltaic system. These customers will draw less load from the main electric grid when they are self-generating.

Many facets of modern society are heavily reliant on the main electric grid, and a major outage for an extended duration can have severe consequences. Several other categories of infrastructure, including water, transportation, and communications are heavily dependent on electric power infrastructure. Services including healthcare, emergency operations, command and control centers, municipal services, wastewater treatment plants, data centers, banking, and more can be affected by a loss of electric power. [5]

Many critical facilities have individual building-tied backup generators and uninterruptible power supplies (UPS) to maintain critical loads for a short duration blackout of the main electric grid. However, these resources often have not been designed or maintained to support longer-term outages from expanding types and levels of threats and disruptions. Natural disasters such as hurricanes, floods, and tornadoes, as well as intentional attacks such as cyber or physical attacks to grid infrastructure can cause outages lasting for weeks or more. Stored fuel for generators typically lasts only a few days without external refueling from central storage sites; sites which may also be affected by the event causing the extended electric grid outage. Because of the interdependency of critical services, a loss of power in one location can adversely affect other functions or operations at other locations, potentially leading to a chain of events that could have a devastating impact on overall critical services.

2.2.2 Functional Categories of Minigrids

A simple definition of a minigrid is a set of loads with local generation that can be isolated from the main electric grid. As seen in Figure 11, minigrids can be single customer solutions, may serve several customers as a partial feeder minigrid, or may encompass a full feeder or substation.

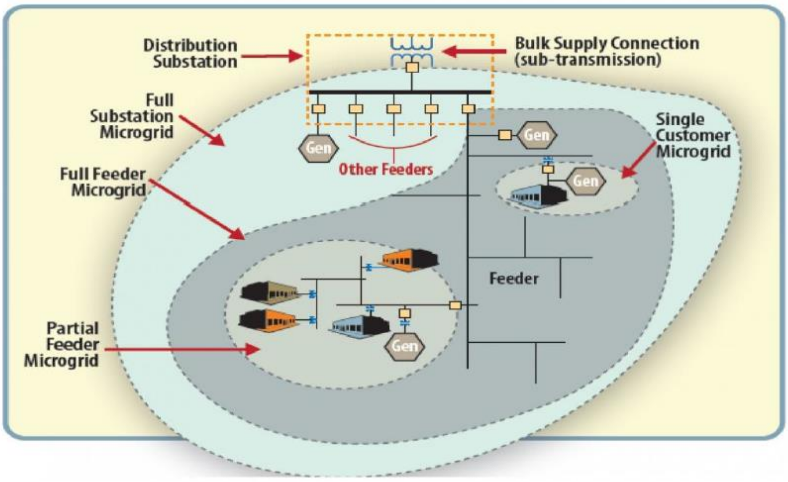

Fig. 11: Illustration depicting the various possible sizes of minigrids (Image from: https://www.energy.gov/oe/activities/technology-development/grid-modernization-and-smart-grid/role-microgrids-helping)

The basic operation of a minigrid can be separated into three main types based on (a) whether the minigrid is typically connected to the main electric grid or typically islanded and (b) if the minigrid has enough generation for sustained operation or simply short-term backup generation.

We separate minigrids into three basic types:

- Type 1: Minigrid for Backup Only
  - Operates only when the main electric grid is down
  - Generation is sized to cover critical loads only
- Type 2: Always Islanded Minigrid
  - Never connected to the main electric grid (e.g., a remote system far from the main grid)
  - Has enough local generation to cover all local load

- Type 3: Hybrid Minigrid
    - Operates grid-connected part of the time and islanded part of the time
        - Operation mode determined by factors including costs, main grid outages, fuel supplies, etc.
    - Has enough local generation to cover all local critical loads, may have enough generation to cover all local loads

Type 1 minigrids provide backup power to critical buildings when utility power is lost by opening the point of common coupling (PCC) main breaker switch, isolating the system form the main grid. After isolation, there is startup and synchronization of generators to the critical loads served.

While the simplest Type 1 minigrid would be one generator and one critical load, the most effective Type 1 minigrids involve multiple generators and multiple critical loads, because additional generators provide redundancy and coordinated controls will make the generators run more efficiently, resulting in efficient, reliable, and resilient backup power.

Type 2 minigrids involve simply local generation and load and are never connected to the main grid. These systems may be referred to as “off-grid.” In Type 2 minigrids, it is essential to appropriately match generation and load for continuous operation. Type 2 minigrids will require larger generation resources, fuel supplies, and energy storage systems than Type 1 or Type 3, since they must constantly operate autonomously. Although there is no switch needed for isolation from the main electric grid, Type 2 may have isolation switches to separate critical loads from non-critical loads during periods of low generation (e.g., due to a fuel shortage or a lack of wind or solar resource).

Type 3 minigrids are the most flexible option. These minigrids can operate either grid-tied or islanded from the main electric grid. Type 3 minigrids will at least have generation to cover their critical loads, and often will have generation to cover all loads. The later scenario of generation to cover all loads provides significant flexibility to respond to grid signals such as time of use pricing, demand response requests, or grid outages while maintaining reliable power for all loads on the minigrid. During times of high minigrid load, the minigrid may draw power from the main electric grid to supplement its local generation. During times of low minigrid load, it may be possible to sell power back to the main grid. Sending power back to the main grid may be particularly valuable during periods of main grid peak load and during resilience events which stress the main grid.

Generation resources on minigrids are distributed energy resources (DERs). DERs can include diesel and gas engines, microturbines, fuel cells, PV, wind, biomass, and energy storage. These local generation resources enhance reliability by providing power to the minigrid's critical resources when the minigrid is islanded. When not islanded, excess generation may be able to be sold back to the utility to offset DER capital and operation costs.

DERs can also be used as peak shaving devices, operating only when the minigrid loads are large and it is desired to reduce net consumption from the utility (e.g., to minimize a capacity cost).

Site requirements will impact which generation resources are best and how the generators are able to run. For example, United States Environmental Protection Agency standards limit both NOX emissions for diesel engines and the number of hours that diesel engines can run, which can limit their ability to supply power to serve loads except under emergency conditions – and may make diesel-only systems most appropriate for Type 1 backup-only minigrids. Renewable energy including solar and wind power, especially when paired with energy storage, is particularly attractive for Type 2 and Type 3 minigrids, though wind and solar resources vary by location and season.

In many cases, a large amount of generation needs can be supplied by the renewable resources and supplemented as needed by other generation such as diesel generators or by drawing power from the main grid for Type 3 minigrids.

A minigrid should have capabilities designed to make it operate with flexibility and efficiency. [7]

Some important capabilities include:

- Flexibility in placement and technologies associated with generation resources including distributed generation, renewables, and energy storage by development of plug-and-play capabilities. Plug-and-play also provides for reduction of engineering costs and increased reliability through shared use among multiple facilities within the microgrid. There may be a range of different sizes of generation resources in the microgrid.
- Complex controls including dynamic power quality control, intentional islanding, and autonomous control of generation resources. These complex controls allow the minigrid to provide high-quality power efficiently even when not connected to the main electric grid.
- System robustness through the ability of generation resources to coordinate to meet the needs of the loads. The minigrid provides for continuous operation during loss of the utility grid and compensates for loss of local generation resources by sharing loads between units.
- Efficient operations by matching total generation to the minigrid load (with a slight excess for contingencies), the generation resources are run more efficiently so only the backup generation required for the minigrid is utilized.

Minigrids are designed to distribute existing and new generation resources among buildings to meet critical energy needs.

Minigrid implementation may require the following types of alterations to typical infrastructure associated with drawing power from the main electric grid:

- Additional transformers/breakers/controls to existing generator resources (backup generators, PV, etc.) – step up voltage levels of backup generators to designated feeder levels, if necessary, and apply minigrid monitoring and generator resource controls of voltage and power levels
- New generation resources (generators, PV, etc.) – add sufficient new generation resources to supply required critical minigrid load demand when the minigrid is islanded from the utility grid, assuming minigrids have enough generation such that the loss of any generation resource within the minigrid will not entail loss of load (which provides so-called 'N-1' redundancy)
- Static switch/main breaker – provide a main isolation device separating the minigrid from the main electric grid to allow it to change between grid-tied and islanded (note: there may be multiple isolation devices between a minigrid and the utility grid)
- Sectionalizing switches/breakers – can be used to isolate non-critical loads within a minigrid when limited generation is available to serve loads or to sectionalize a minigrid into zones of protection to isolate faults
- Energy storage – protect non-interruptible loads and provide ride-through capability until distributed generators start up; can also improve system performance, such as absorbing sudden changes in PV, so that generators limit the amount of ramping in response to PV fluctuations
- Minigrid controls – use a set of centralized and distributed controls to monitor and control generation resources or isolation devices (breakers, switches) to switch the minigrid between grid-tied and islanded operation, as well as deploy the generator resources efficiently to reduce fuel use by being responsive to load conditions
- Protection – minigrid system protection against fault conditions to isolate generation devices from the system during the operation
- Building load reconfiguration – in some minigrid designs, the critical load needs for a minigrid can be reduced by reconfiguring building loads to sectionalize critical and non-critical loads within the building so that the minigrid is only required to supply a portion of building loads rather than entire building loads
- Load shedding – in some minigrid designs, isolation devices can isolate less critical loads within a minigrid when sufficient generation is not available to meet all the load within the minigrid
- New feeders – in some minigrid designs, it may be more economical to install a new dedicated minigrid feeder connecting critical buildings together rather than

use the existing utility grid because the amount of non-critical load far exceeds the critical load (so it would be cost prohibitive to use the existing utility grid to form a minigrid)

- Feeder rearrangement – in some designs, instead of installing a new dedicated feeder, it may be possible to reconfigure the connections of an existing utility feeder so that critical loads are on the minigrid feeder and the non-critical loads are on other feeders (this existing feeder can be made into a minigrid without a prohibitively large amount of generation required to meet loads).

Energy storage with fast response times can be used to keep non-interruptible loads from experiencing short outages during a minigrid's transition between grid-tied and islanded mode.

Without energy storage, there may be a short outage (e.g., 10 - 60 seconds) when transitioning from grid-tied to islanded as microgrid generation resources start up and synchronize to a standard frequency. Non-interruptible critical loads, such as telecom or computer server equipment, are usually equipped with uninterruptible power supply (UPS) units to provide five or more minutes of backup power to these loads. The power is rated to ride through the time necessary for backup diesel generators to start and recharge the batteries. A minigrid could be designed to allow ride-through of all critical loads by using many UPS units, but if an entire building requires non-interruptible loads, then a larger scale energy storage unit may be most effective.

Energy storage has additional benefits of being able to help control variation in generation. The storage system can dampen the variability of solar or wind systems caused by cloud cover changes or shifts in wind. If large enough, energy storage may also be able to help address daily variability such as evening peaks in load and the diurnal cycle of PV power (i.e., no solar irradiance at night). As a rough rule of thumb, it has often been cost-effective to install some amount of energy storage when variable generation exceeds about 20% of total minigrid generation to prevent excessive ramping of other generation resources (diesel or natural gas generators, microturbines, etc.). Engineering studies considering renewable variability, cost, and the system's other generators' performance will inform the optimal balance of renewable resources with energy storage.

Building load reconfiguration refers to how the existing emergency connections of critical buildings are setup and what adjustments can be made to prioritize critical loads. Buildings with backup generation generally have an automatic transfer switch (ATS) that closes the generator onto a portion of the building loads during emergency situations. If it is determined that a larger portion of a critical building should be supplied by the minigrid, then existing switchboards and/or panelboards will have to be retrofitted or expanded to accommodate the new load requirements. Or, if a new building is added to a minigrid, it might be desired to reconfigure the building so only the critical loads in the building are connected to the minigrid to limit the amount of generation required on the minigrid. [9]

Non-critical loads can be shed by installing remotely operable main breakers on the incoming building feeds, which will isolate these buildings when the minigrid is in islanded mode. If the minigrid is designed to handle all loads within its jurisdiction, these retrofits won't be required, but additional generation will be required to cover these additional loads.

If it is too cumbersome to create a minigrid within an existing distribution feeder system, it may be possible to reroute a portion of the non-critical loads along the existing radial distribution feeder to other feeders.

This will allow the minigrid to island from the utility during power outages and supply mostly critical loads so that generation requirements are reduced. It also may be more efficient to develop a separate dedicated minigrid feeder that is attached to only critical loads, isolated from the utility by one or more PCCs, to reduce the amount of generation required for the minigrid.

2.2.3 Performance Risk Analysis

It is important to evaluate the ability of the energy system to meet the defined extended outage performance criteria. This is typically done using a risk-informed performance assessment.[9] Described in this section are simple performance parameters used to defined performance risk in a way that has been valuable to previous analyses. However, this definition of performance risk may need to be modified based on the specifics of the minigrid being considered to best address the true performance of that microgrid. For example, the equations below reference percent of critical buildings served. For certain applications, the percent of people receiving a critical service (e.g., clean water, cell phone signal, etc.) may be a better metric.

We have generally based energy system performance risk assessment on how well the energy system can meet critical infrastructure functions and services during a given power outage. Based on this approach, we define the performance risk PR for a given outage as a function of the critical buildings and loads served and the length of time they can be met by the energy system. The performance risk, PR, defined as:

$$PR = 1 - \left(CBS * CLS * RG * \frac{Da}{Dn} \right) \qquad (2)$$

Where:

$CBS =$ Percent of critical building served – critical buildings with backup power systems. If few buildings are served, then consequences and risk will be high.

$CLS =$ Percent of critical loads served – weights serving the defined critical loads for the critical services and buildings. If minimal loads are covered, the consequences and risks will be high.

$RG =$ Reliability of generation – weights the maintenance of backup generators. Low maintenance lower reliability and the risks will be high.

$\frac{Da}{Dn} =$ Ration of generator fuel availability versus outage duration. If the generator fuel tank is small, and/or the ability to refuel the generator is low, then the risks can increase for longer power outages, unless renewable or other energy resources are available.

Based on customer outage evaluations for some major natural disasters, it has been found that typically when backup power systems can meet 85% or more of the critical buildings and loads served for 85% or more of the outage duration, the overall power system can adequately provide power to support critical community services and functions without significantly impacting overall public health and safety. For energy systems that meet less than 70% of the critical buildings and loads served for less than 70% of the outage duration, the community health and safety become increasingly stressed. Therefore, in general we have quantified energy system performance risk notionally as:

Low Performance Risk – $PR < 0.30$

Medium Performance Risk – $0.30 < PR < 0.50$

High Performance Risk – $PR > 0.50$

2.2.4 Resilience Enhancements to Improve Performance

It has been evaluated improved resilience as both the reduced impact of the event and the reduced recovery time to return to normal operation after the event. Specifically, the system impact (SI) of the event is the time integral of the "typical" system performance (TSP) minus the actual system performance (SP):

$$SI = \int_{t_0}^{t_f} [TSP(t) - SP(t)]dt \qquad (3)$$

Similarly, the total recovery effort (TRE) is the time integral of the recovery effort:

$$TRE = \int_{t_0}^{t_f} [RE(t)]dt \qquad (4)$$

Improved resilience will minimize both the system impact and the total recovery effort, as illustrated in Figure 12:

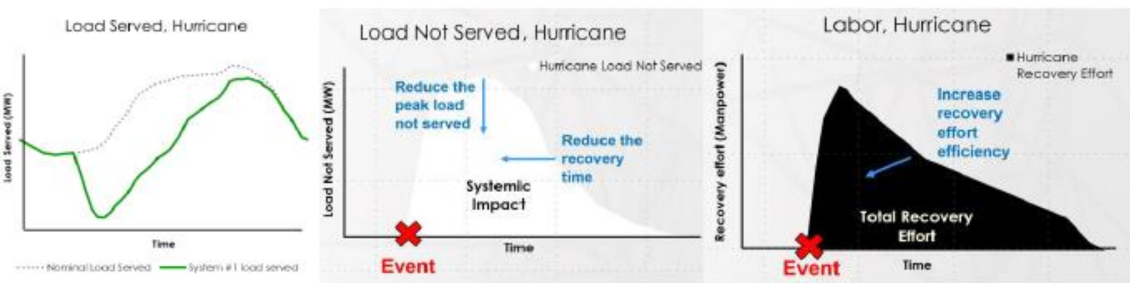

Fig. 12: Hypothetical impact of a hurricane, showing (left) normal load and actual load served, (center) load not served, and (right) labor required for recovery. Blue arrows and text indicate the goals of resilience to reduce system impact and total recovery effort.

2.2.5 Formulating and Evaluating Design Options

The main goal is to formulate design options based on performance objectives for the set of critical service assets required to serve during the Design Basis Threat (DBT) event. To do this, I utilize methods and tools to come up with a set of resilient design options. Part of this analysis may be to cluster critical assets and to overlay these clusters onto the existing distribution system to determine which areas might be initial microgrid candidates. One can then use performance metrics to further define and select which of the initially identified candidates should be further developed with conceptual microgrid designs for resilience improvements and what additional assets might require hardening for resilience even if a microgrid is not implemented.

The *Initial Conceptual Design Phase* (10-15% design) is focused on the development of initial project scope, objectives, and requirements.

This provides a general description of the major design and construction elements, best locations of minigrid components to enhance energy surety, and suggestions of the elements and operational scenarios to be included. A flow chart of the initial design process is seen in Figure 13.

The process begins with a vulnerability analysis study to determine parts of the system most likely to be impacted by the events described in the Design Basis Threat (DBT) and for which minigrids might be of most value. For example, communities connected by overhead power lines and on the end of a distribution feeder may be especially vulnerable and hence especially good candidates for minigrids, as failures anywhere along the feeder could cut off their connection to the main electric grid. The design options identified for consideration will represent a set of options that may improve the surety of the system for the critical loads, DBTs, and performance metrics that were identified in previous modules.

Once the design options are identified, quantitative evaluation of the system-level impact of the proposed design options is done through simple simulation of system performance. implemented.

The *Final Design Phase* (30% design) considers the several designs evaluated in the initial design phase and selects a final conceptual design. The initial conceptual design renders several options for meeting the same set of surety goals by either using different technologies or deploying similar technologies in different manners. The final conceptual design takes those initial conceptual designs, expands them using more accurate models/descriptions, and performs detailed studies to determine which option should be implemented based on factors including feasibility, cost, and performance.

Technical feasibility is evaluated in detail during the final conceptual design using steady-state and dynamic simulation and optimization tools.

There are two cost aspects considered: capital costs and operating costs. These costs are studied in detail using capital and installation cost estimates for each option, and simulation of daily, weekly, and seasonal operations under different system conditions to account for the variation in inputs such as renewable generation, fuel costs, and loads.

The performance of the system will be measured in terms of the energy surety goals and the project scope. For example, if increased reliability is a focus, then performance can be measured in terms of improvement in reliability metrics such as System Average Interruption Duration Index (SAIDI), System Average Interruption Frequency Index (SAIFI), etc. [9]

Detailed schematics will also be developed during the final design phase and will be shared with the engineering firm that will ultimately be responsible for constructing the minigrid.

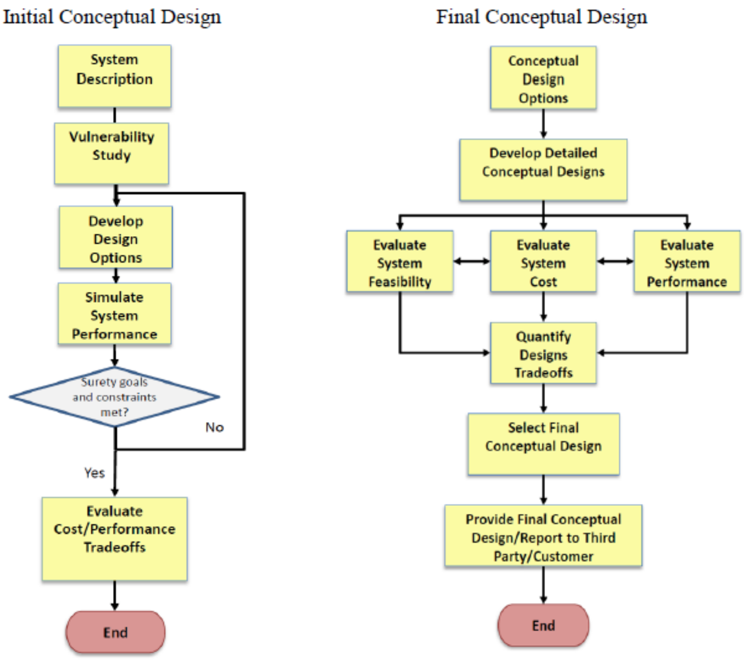

Fig.13: Initial (left) and final (right) conceptual design process.

2.3 Resilience Engineering: Fundamental Concepts

Resilience engineering is a relatively new field of research that has gained significant attention in recent years. It is an interdisciplinary field that draws on concepts and methods from various disciplines such as systems engineering, risk management, and organizational studies. Resilience engineering is concerned with the ability of complex systems to withstand and recover from disruptions, shocks, and failures. It aims to understand how organizations and systems can be designed and managed to be more resilient in the face of uncertainty and dynamic environments. [6]

The study of resilience engineering has become increasingly important in today's complex and rapidly changing world. [11] As systems become more interconnected and interdependent, the likelihood of disruptions and failures increases. Resilience engineering provides a framework for understanding and addressing these challenges by focusing on the design and operation of systems that can absorb and adapt to disruptions while maintaining their functionality and performance.

One of the key concepts in resilience engineering is the idea of resilience thinking. This approach emphasizes the importance of understanding the dynamics of complex systems and the interdependencies between their components. Resilience thinking involves anticipating and preparing for potential disruptions, monitoring and responding to threats in real-time, and adapting to changing conditions to maintain system functionality. [12]

Another important concept in resilience engineering is the concept of the "resilience triangle" (see Figure 14). This framework identifies three critical components of resilience: robustness, redundancy, and response. Robustness refers to the ability of a system to withstand disturbances without collapsing. Redundancy refers to the presence of backup systems or resources that can take over in case of failure. Response refers to the ability of the system to quickly adapt and recover from disruptions.

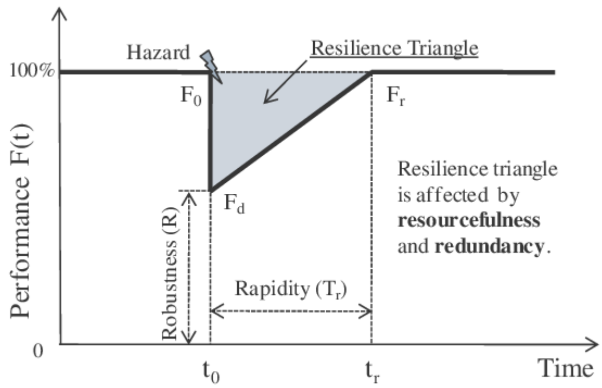

Fig.14: Graphical depiction of resilience triangle and resilience four attributes (Y.Zhang & al)

Resilience engineering has applications across a wide range of domains, including aviation, healthcare, finance, and critical infrastructure. For example, in aviation, resilience engineering has been used to develop systems that can withstand unexpected events such as equipment failure or severe weather conditions. In healthcare, resilience engineering has been used to design systems that can respond to unexpected patient surges or infectious disease outbreaks.

The development of resilience engineering has been influenced by various disciplines, including systems engineering, risk management, and organizational studies. Systems engineering provides a framework for understanding the design and operation of complex systems, while risk management provides tools for identifying and assessing potential threats. Organizational studies provide insights into the role of human factors and organizational culture in shaping resilience.

Resilience engineering has also been influenced by advances in technology, particularly in the field of data analytics. The use of machine learning and artificial intelligence has enabled organizations to detect and respond to potential threats in real-time. For example, predictive analytics can identify patterns of behavior that may indicate a potential security threat, while machine learning algorithms can help organizations optimize their response to disruptions. [15]

Despite the progress made in resilience engineering, there are still significant challenges that need to be addressed. One of the main challenges is the need to balance resilience with efficiency and cost-effectiveness. Resilience engineering often requires investing in redundant systems, developing contingency plans, and implementing new technologies, which can be costly. Another challenge is the need to address the human factor in resilience engineering. Organizations need to create a culture that values resilience and encourages employees to speak up and report potential threats.

Resilience engineering is a crucial field of research that has significant implications for organizations and society as a whole. As systems become more complex and interconnected, the need for resilience engineering will only continue to grow. By understanding the principles of resilience engineering, organizations can design and operate systems that are better equipped to withstand and recover from disruptions, ultimately reducing the risk of catastrophic failures and improving overall performance. Furthermore, resilience engineering has the potential to enhance the sustainability and adaptability of systems, which is critical in today's rapidly changing world.

2.3.1 Hazard types and impact on community-level energy systems

As shown in Figure 15, meteorological events, hydrological events, geographical events and climatological events have been increasing throughout the world. The magnitude and frequency of extreme events are expected to increase even further in the following years. Understanding the hazards is crucial for resilient design and preparation against extreme events. Although various hazards may manifest differently, they can all be categorised based on their impact, frequency, return period, geographical probability, event duration, and warning time. In Figure 15, hazards are categorised based on their duration and warning time. Acute threats include sudden hazards such as hurricane, tornado, bushfire, earthquake, pandemic or cyber-attacks; and chronic stresses include slow and mostly cyclical hazards such as drought, chronic flooding, sea-level rise and increases in ambient temperature.

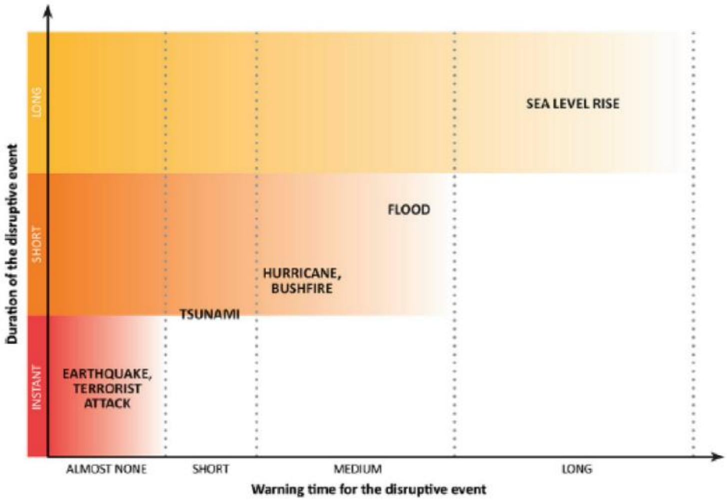

Fig.15: Various types of hazards based on their duration and warning time obtained from [5].

The disruption events can be categorised based on the impacts on the energy entity that results in declining availability, affordability, or acceptability of energy metrics. Under some circumstances, the combination of extreme events (multiple hazards) might create a high impact threat with distinct characteristics. Extreme events can lead to prolonged and severe failures that threaten lives, political landscapes, and businesses. Therefore, these threats should be identified and their potential impact on energy system components should be investigated. [14]

Energy systems are formed from interdependent components and may be weather-dependent. The current energy systems have been designed based on previous climate-related assumptions, where climate-induced energy supply and demand variations were based on historical assumptions.

The impacts of increasing intensity and frequency of extreme events caused by climate change on different components of the energy supply chain are discussed in several studies [14 & 15].

The generic short and long-term impacts of the typical climatic hazards on various aspects of energy systems are summarised in Table 5. However, the quantification of these impacts is challenging due to high uncertainties in the models and inputs [16]. Community-level energy systems are a part of the national and regional energy systems. Often, the impact of the extreme events on the larger energy system also has inevitable consequences on the community system. One of the main chronic threats on energy systems today is an increase in ambient temperature due to global warming which directly results in lower efficiency of electricity generation for both conventional fossil fuels and new renewable sources.

Higher temperatures also raise cooling needs and lower the efficiency of mechanical cooling systems (less effective natural ventilation) so in the future there will be an increase in cooling needs, and it is necessary to consider resilience as a factor in the long-term planning and design of buildings. [15]

Items	Short-term Impacts	Long-term impacts
Centralised energy generation (both renewable and non-renewable)	Physical damages cause higher failure/ disruption risk, risk of cascading failure	Efficiency reduction, capacity reduction, higher generation variability, extended disruption period, higher water consumption, relocation of current plants
Local energy generation (e.g. district energy systems and microgrids)	Physical damages cause higher failure/ disruption risk, risk of cascading failure	Efficiency reduction, capacity reduction, high uncertainty in generation forecast, extended disruption period, higher water consumption
Transmission and distribution (e.g. pipelines, poles and wires)	Higher risk of failure/disruption, risk of cascading failure	Capacity reduction, efficiency reduction, accelerated aging, relocation of the current networks, longer disruption period
Energy storage (e.g. batteries, thermal storage and hydro)	Higher risk of failure/disruption	Efficiency reduction, water scarcity may affect investments
Energy use (e.g. commercial and residential uses)	Higher risk of disruption	Increase in peak load, increase in energy demand, change in demand profile, higher demand for air conditioning systems, high uncertainty in demand forecast
Operation and maintenance	Higher risk of operational disruptions, increase in components replacement frequency	High uncertainty in generation forecast, increase in unscheduled maintenance, reduction in capacity factor, new operation strategies
Socioeconomics	Higher uncertainty in investments, higher social burden, increase in physical injuries and trauma	Health risks, higher social and political pressures, technological changes and new markets, cross-sectoral competition (e.g. food, water and energy), increase in energy prices

Tab. 5: Short and long-term impacts of climate-related events on different items of the energy system [5]

In addition to direct impacts, extreme events also have cascading consequences for all stakeholders, such as access to clean water and provision of acceptable indoor air quality. Although estimating these multi-dimensional impacts is not easy, public and private organisations have started to categorise these impacts and employ strategies in the planning and design of energy system components. The goal is for local energy generation in communities to be designed to be more resilient, efficient, and sustainable, e.g. the fourth generation of district heating systems [16], with a low frequency and duration of outages in critical systems. It helps that these systems are located near the energy demands that they serve. Energy delivery distance is usually short and reduces the risks and losses of energy transmission. These systems tend to be more rigorous due to the direct economic impacts on the system owner. However, changes in climate and its associated risk will not only change the planning and design of energy infrastructures but also their operation and maintenance.

In energy master planning at the community level; such as in a municipal district heating and cooling system, an isolated indigenous community, or a hospital campus; the impacts of disruptive climate-related events should be considered in early stages of the design and planning of the location’s energy system to increase resilience.

2.3.2 Robustness and Reliability

The term robustness can sometimes be confused with resilience but a more explicit definition of the former is "the ability of the system to withstand a given level (of disruption)" [12]. The Department of Energy (DOE) in the US defines reliability in the electricity grid context as "the ability of the system or its components to withstand instability, uncontrolled events, cascading failures, or unanticipated loss of system components" [12].

Reliability focuses on energy systems being able to provide service during disruptions which occur frequently. A reliable system and system components offer long-term and robust performance in their intended function. Reliability can be defined both in terms of system performance and individual system component performance. The focus here is on system level robustness and reliability. The reliability of systems is usually expressed as a coefficient in design, which usually covers short-term measures such as emergency systems and energy reduction measures. For example, In Australia, the reliability standard in the electrical system is equal to %0.002 of the unserved energy of the total energy demand in each region per year [15] for all expected events. The robustness and reliability of the system are pre-disruption characteristics of a system. However, climate change and the transition to clean energy systems limit the robustness and reliability of the systems. The changing conditions require a dynamic response to the combined impacts of climate risks with both short and long-term planning and enhancement measures. Although a variety of scenarios are considered in estimating the reliability of the future energy systems the response to extreme climate events seems to be lacking in these scenarios.

Reliability and resilience have different characteristics in a power system context. These concepts need to be harmonised with the broader engineering literature and other industry sectors. In addition, the evolving risks require more explicit metrics, as well as comprehensive approaches, policies, and regulations that can guarantee the resilience of the energy system and its ability to provide valuable services to communities and other industry sectors.

2.3.3 Energy Resilience

Although there are various definitions of system resilience in the literature, the resilience of modern composite energy systems requires more research, especially at the community level including the new generation of district and smart energy systems. [14]

The resilience of energy systems can be defined based on the characteristics of these systems and the nature of the disruptive events (e.g. a system can be resilient to a heatwave but not to an ice storm). For example, DOE defines the resilience of the electrical grid as "the ability of a system or its components to adapt to changing conditions and withstand and rapidly recover from disruptions" (opposed to "security" which is to "withstand attacks"). There are six components of resilience, namely "the ability of an entity - asset, organisation, community, region - to anticipate, resist, absorb, respond to, adapt to, and recover from a disturbance" [5].

A sustainable energy system (i.e. satisfying four dimensions of availability, accessibility, affordability and acceptability) should be comprised of preparation, absorption, recovery and adaptation abilities. In order to provide more resilience to increasingly complex and interconnected (energy) systems and tackle the uncertainties, costs, and challenges in the nature of these systems subject to extreme events need for clear definitions, metrics, and evaluation methods for resilience development.

Furthermore, the performance criteria that include these metrics need to be formulated more in probabilistic terms to account for risks and likelihoods of disruption. In the Sandia National Laboratory report for the 2015 Quadrennial Energy Review, the authors note that resilience metrics should consider threat, likelihood, and consequence and thus because common reliability metrics do not possess these attributes, they are "orthogonal in purpose and discrimination capability to resilience metrics" [9]. Resilience metrics encompass all disruptions that have different levels of uncertainties, with particular attentiveness to high-impact, low-frequency events [15]. But in addressing system resilience, both qualitative and quantitative performance criteria and metrics are required.

Quantitative, time-dependent resilience metrics are introduced for power system resilience to measure how fast and how low the resilience drops, how long the system remains in the degraded state, and how quickly it recovers.

Although these metrics from the literature are not focused solely on energy resilience, they can be tailored for community-level system energy resilience evaluation. A summary of resilience metrics, and qualitative and quantitative resilience evaluation methods, is given in Table 6.

Type	Metrics	Evaluation Method
Qualitative Evaluation	- Resiliency indices - Functional redundancy	- Checklists and questionnaires - Matrix scoring system - Analytic hierarchy process (AHP) - Energy flow-based system performance modelling methods under different scenarios - Graph-theory and probabilistic method - Spatial power outage duration model - Benefit-cost analysis
Quantitative Evaluation	- Time-dependent metrics for the resilience of power networks based on slopes and area of resilience trapezoid - Probability distribution of economics costs - Area under the curve between targeted performance and real performance - Ratio of the area between real performance curve to targeted performance curve during a year - Probability of network performing its intended functions - Time to restoration following a failure - Performance-based resilience index	

Table 6: Resilience metrics and evaluation methods

It can be seen that although there are several metrics and evaluation methods for specific events, there is no consensus on method or metrics for measuring energy resilience, and defining the mitigation and enhancement strategies, especially during the energy master planning. Various aspects of resilience are considered individually and with a narrow focus which restricts opportunities in other dimensions of resilience e.g. infrastructure, operational and social. Although these qualitative and quantitative methods and metrics can lead to measurement of a particular resilience measure, they do not provide a consistent approach for measuring the overall resilience of the system for the purpose of energy resilience planning.

In energy master planning process, prescriptive and performance approaches (and associated metrics) can be employed in resilience assessment and enhancement. Prescriptive-based resilience approach considers the acceptable or required resilience solutions and limits, while the performance-based resilience approaches use quantifiable metrics to measure the resilience of the system performance. The performance-based approach includes the system-based or attribute-based metrics such as level of redundancy, number of backup transformers and number of highly trained staff. Resilience planning for critical infrastructure should be based on the critical services required to support the community rather than the physical condition of infrastructure only [14].

The system performance metrics can be divided into the consequence-based (e.g. environmental, social, economic and the national security) metrics and service-based (e.g. electricity, heating, cooling and water) metrics. Sometimes the resilience performance of each service should be measured in each critical nodes of the energy system. Each of these nodes might have different resilience requirements.

The performance-based metrics can be employed in energy master planning to estimate the impacts of the system disruption or unusual system service performance in terms of environmental, social and economic consequences. These consequences are sometimes inter-related for instance, assessing the social impacts might be required to be able to assess the economic impacts.

2.4 EDA Exploratory Data Analysis

Exploratory Data Analysis (EDA) refers to the method of studying and exploring record sets to apprehend their predominant traits, discover patterns, locate outliers, and identify relationships between variables. EDA is normally carried out as a preliminary step before undertaking extra formal statistical analyses or modeling. [17]

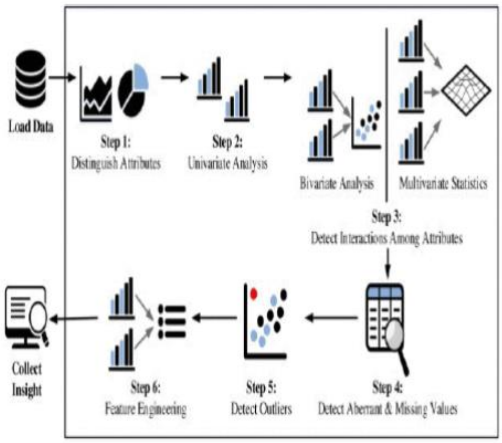

Fig. 16: Different stages of Exploratory Data Analysis [19]

2.4.1 EDA breakdown

A logical step-by-step breakdown for EDA can be summarised as: [18,19]

1. Data Cleaning: EDA involves examining the information for errors, lacking values, and inconsistencies. It includes techniques including records imputation, managing missing statistics, and figuring out and getting rid of outliers.

2. Descriptive Statistics: EDA utilizes precise records to recognize the important tendency, variability, and distribution of variables. Measures like suggest, median, mode, preferred deviation, range, and percentiles are usually used.

3. Data Visualization: EDA employs visual techniques to represent the statistics graphically. Visualizations consisting of histograms, box plots, scatter plots, line plots, heatmaps, and bar charts assist in identifying styles, trends, and relationships within the facts.

4. Feature Engineering: EDA allows for the exploration of various variables and their adjustments to create new functions or derive meaningful insights. Feature engineering can contain scaling, normalization, binning, encoding express variables, and creating interplay or derived variables.

5. Correlation and Relationships: EDA allows discover relationships and dependencies between variables. Techniques such as correlation analysis, scatter plots, and pass-tabulations offer insights into the power and direction of relationships between variables.

6. Data Segmentation: EDA can contain dividing the information into significant segments based totally on sure standards or traits. This segmentation allows advantage insights into unique subgroups inside the information and might cause extra focused analysis.

7. Hypothesis Generation: EDA aids in generating hypotheses or studies questions based totally on the preliminary exploration of the data. It facilitates form the inspiration for in addition evaluation and model building.

8. Data Quality Assessment: EDA permits for assessing the nice and reliability of the information. It involves checking for records integrity, consistency, and accuracy to make certain the information is suitable for analysis.

2.4.2 Types of EDA

EDA, or Exploratory Data Analysis, refers back to the method of analyzing and analyzing information units to uncover styles, pick out relationships, and gain insights. There are various sorts of EDA strategies that can be hired relying on the nature of the records and the desires of the evaluation. Here are some not unusual kinds of EDA: [18,19]

1. Univariate Analysis: This sort of evaluation makes a speciality of analyzing character variables inside the records set. It involves summarizing and visualizing a unmarried variable at a time to understand its distribution, relevant tendency, unfold, and different applicable records. Techniques like histograms, field plots, bar charts, and precis information are generally used in univariate analysis.

2. Bivariate Analysis: Bivariate evaluation involves exploring the connection between variables. It enables find associations, correlations, and dependencies between pairs of variables. Scatter plots, line plots, correlation matrices, and move-tabulation are generally used strategies in bivariate analysis.

3. Multivariate Analysis: Multivariate analysis extends bivariate evaluation to encompass greater than variables. It ambitions to apprehend the complex interactions and dependencies among more than one variables in a records set. Techniques inclusive of heatmaps, parallel coordinates, aspect analysis, and primary component analysis (PCA) are used for multivariate analysis.

4. Time Series Analysis: This type of analysis is mainly applied to statistics sets that have a temporal component. Time collection evaluation entails inspecting and modeling styles, traits, and seasonality inside the statistics through the years. Techniques like line plots, autocorrelation analysis, transferring averages, and ARIMA (AutoRegressive Integrated Moving Average) fashions are generally utilized in time series analysis.

5. Missing Data Analysis: Missing information is a not unusual issue in datasets, and it may impact the reliability and validity of the evaluation. Missing statistics analysis includes figuring out missing values, know-how the patterns of missingness, and using suitable techniques to deal with missing data. Techniques along with lacking facts styles, imputation strategies, and sensitivity evaluation are employed in lacking facts evaluation.

6. Outlier Analysis: Outliers are statistics factors that drastically deviate from the general sample of the facts. Outlier analysis includes identifying and knowledge the presence of outliers, their capability reasons, and their impact at the analysis. Techniques along with box plots, scatter plots, z-rankings, and clustering algorithms are used for outlier evaluation.

7. Data Visualization: Data visualization is a critical factor of EDA that entails creating visible representations of the statistics to facilitate understanding and exploration. Various visualization techniques, inclusive of bar charts, histograms, scatter plots, line plots, heatmaps, and interactive dashboards, are used to represent exclusive kinds of statistics.

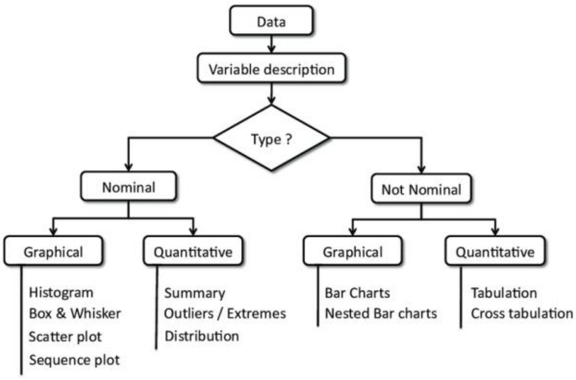

Fig.17: General overview of Exploratory Data Analysis (F. Desmet & al)

These are just a few examples of the types of EDA techniques that can be employed at some stage in information evaluation. The choice of strategies relies upon on the information traits, research questions, and the insights sought from the analysis.

2.5 Anomaly Detection

2.5.1 Anomaly Detection Framework

Anomaly detection is the process of identifying data points, entities or events that fall outside the normal range. An anomaly is anything that deviates from what is standard or expected. Humans and animals do this habitually when they spot a ripe fruit in a tree or a rustle in the grass that stands out from the background and could represent an opportunity or threat. Thus, the concept is sometimes framed as outlier detection or novelty detection. [20]

Anomaly detection has a long history in statistics, driven by analysts and scientists who pored over charts to find elements that stood out. Over the last several decades, researchers have started automating this process using machine learning training techniques designed to find more efficient ways to detect different types of outliers.

In practice, anomaly detection is often used to detect suspicious events, unexpected opportunities or bad data buried in time series data. A suspicious event might indicate a network breach, fraud, crime, disease or faulty equipment. An unexpected opportunity could involve finding a store, product or salesperson that's performing much better than others and should be investigated for insight into improving the business. [21]

An anomaly could also be the result of faulty equipment, broken sensors or a disconnected network. In these instances, a data scientist might want to remove the anomalous data records from further analysis so as not to compromise the development of new algorithms.

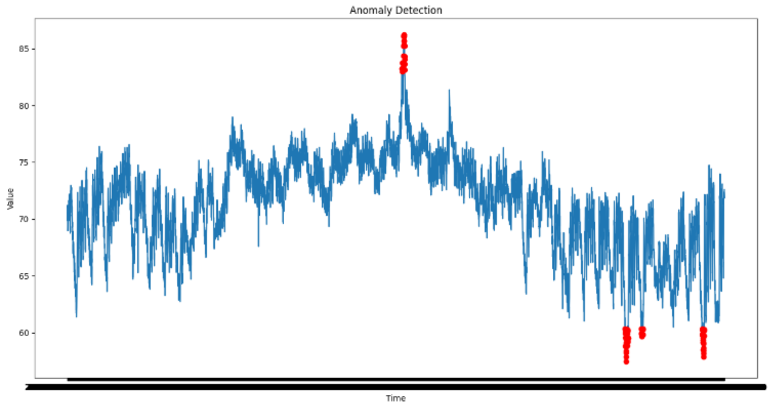

Fig. 18: Anomaly Detection in Time Series Data (https://www.geeksforgeeks.org/anomaly-detection-in-time-series-data/)

There are several ways of training machine learning algorithms to detect anomalies. Supervised machine learning techniques are used when you have a labeled data set indicating normal vs. abnormal conditions. For example, a bank or credit card company can develop a process for labeling fraudulent credit card transactions after those transactions have been reported. Medical researchers might similarly label images or data sets indicative of future disease diagnosis. In such instances, supervised machine learning models can be trained to detect these known anomalies.

Researchers might start with some previously discovered outliers but suspect that other anomalies also exist. In the scenario of fraudulent credit card transactions, consumers might fail to report suspicious transactions with innocuous-sounding names and of a small value. A data scientist might use reports that include these types of fraudulent transactions to automatically label other like transactions as fraud, using semi-supervised machine learning techniques. The supervised and semi-supervised techniques can only detect known anomalies. However, the vast majority of data is unlabeled. In these cases, data scientists might use unsupervised anomaly detection techniques, which can automatically identify exceptional or rare events. [20]

For example, a cloud cost estimator might look for unusual upticks in data egress charges or processing costs that could be caused by a poorly written algorithm. Similarly, an intrusion detection algorithm might look for novel network traffic patterns or a rise in authentication requests. In both cases, unsupervised machine learning techniques might be used to identify data points indicating things that are well outside the range of normal behavior. In contrast, supervised techniques would have to be explicitly trained using examples of previously known deviant behavior.

Broadly speaking, there are three different types of anomalies.

- Global outliers, or point anomalies, occur far outside the range of the rest of a data set.
- Contextual outliers deviate from other points in the same context, e.g., holiday or weekend sales.
- Collective outliers occur when a range of different types of data vary when considered together, for example, ice cream sales and temperature spikes.

Many different kinds of machine learning algorithms can be trained to detect anomalies. [21] Some of the most popular anomaly detection methods include the following:

- Density-based algorithms determine when an outlier differs from a larger, hence denser normal data set, using algorithms like K-nearest neighbor and Isolation Forest.
- Cluster-based algorithms evaluate how any point differs from clusters of related data using techniques like K-means cluster analysis.
- Bayesian-network algorithms develop models for estimating the probability that events will occur based on related data and then identifying significant deviations from these predictions.
- Neural network algorithms train a neural network to predict an expected time series and then flag deviations.

Anomaly detection systems can be used in various ways to improve business, IT and application performance. These systems can also enhance the detection of fraud, security incidents and opportunities for innovation. The following are some other common use cases for anomaly detection:

- Predicting equipment failure.
- Detecting early signs of pending IT failures.
- Detection of pricing glitches.
- Enhanced fraud prevention.
- Identifying DDoS attacks.
- Identifying stores and products that do better than expected.
- Better product quality.
- Enhanced user experience.
- Cloud cost management.

In cloud cost management, anomaly detection could look for sudden shifts in resource utilization, such as increased database calls, spikes in egress charges or increased SaaS charges. This could help managers identify whether this increase was caused by a new application version release, security breach, or associated with a successful product launch.

In cybersecurity, anomaly detection can evaluate thousands of data streams to detect changes in access requests, an uptick in failed authentications or novel traffic patterns that bear further investigation. Anomaly detection is often built into most security appliances and services for intrusion detection systems, web application firewalls and API security tools. [21]

Application performance management tools commonly monitor logs of all traffic to identify performance issues or failures. In these cases, anomaly detection can allow them to detect new issues not identified with traditional rule-based analysis approaches.

In banking and finance, anomaly detection is commonly used to identify fraud by correlating factors such as the size of transactions, time, location and spending rate. For example, suspiciously large transactions in a foreign country might be flagged. Or a suspiciously large number of smaller transactions from a new vendor might similarly be investigated.

Challenges in anomaly detection [22] include the following:

- Data infrastructure needs to be scaled to support useful anomalies.
- Data quality issues can reduce the performance of anomaly detection.
- Poor anomaly detection algorithms can inundate users with false alerts.

It may take a long time to develop a useful baseline to account for normal patterns like holiday sales, heat waves or other normal things that occur less frequently.

Data scientists, IT managers, security managers and business teams must consider several aspects when designing anomaly detection apps to provide the appropriate value. [23]

- Timeliness. What is the time to value? A fraud detection system must operate in seconds, a security system in minutes, while a business trends analysis app might deliver value with daily updates.
- Scale. Is the objective speed or depth of analysis? Analyzing a few metrics can yield fast results, but deeper insight may require thousands or even millions of data streams.
- Rate of change. How quickly do events being analyzed in the data change? Predictive maintenance apps may need to analyze real-time data streams, while business data tends to change more slowly.
- Conciseness. Is there a better way of summarizing insights of interest relevant to decision-makers?
- Defining incidents. How can you automate the process of labeling related types of anomalies to determine root causes and appropriate responses?
- Explainability. Is it enough to determine if an anomalous event has occurred, or should priority be given to algorithms that can explain contributing factors, even if they are not as accurate?

Anomaly detection is generally baked into most modern security, IT management, and fraud detection systems and applications. However, enterprises that want to develop their own anomaly detection algorithms may wish to turn to popular statistics, data science, and mathematical packages and tools. A sampling of popular ones includes the following: [23]

- Anodot, a business monitoring platform that can detect anomalies in business and cloud events.
- Amazon SageMaker, a data science platform that supports anomaly detection.
- ELKI, an open source data mining tool.
- Microsoft AI Anomaly detector service for Azure.
- PyOD, an open source anomaly detection library written in Python.
- Scikit-learn, a popular data science library that supports anomaly detection.
- Wolfram Mathematica, an algorithm development tool that supports anomaly detection.

2.5.2 Anomaly Detection in Off- Grid Photovoltaic Systems

Photovoltaic (PV) systems have gained increasing popularity as a renewable energy source, particularly in off-grid areas where traditional grid connections are not feasible. However, the reliability and efficiency of PV systems rely heavily on the accurate detection of anomalies, which can be challenging due to the variability of solar irradiance, temperature, and other environmental factors. Anomaly detection in PV systems is crucial for identifying issues such as faulty panels, inverter malfunctions, and grid connection failures. This section aims to provide a comprehensive overview of anomaly detection techniques and applications in off-grid PV systems.

The most famous Anomaly Detection Techniques are: [21]

1. Supervised Learning Methods: Supervised learning methods, such as support vector machines (SVM), random forests, and neural networks, have been widely used for anomaly detection in PV systems. These methods rely on historical data to learn patterns and detect deviations from the norm. However, they require a large amount of labeled data, which can be challenging to obtain in practice.

2. Unsupervised Learning Methods: Unsupervised learning methods, such as k-means clustering, principal component analysis (PCA), and one-class SVM, are used to identify patterns and anomalies in PV system data without the need for labeled data. These methods are particularly useful for detecting novel anomalies that may not have been observed during the training process.

3. Hybrid Methods: Hybrid methods, which combine supervised and unsupervised learning techniques, have also been proposed for anomaly detection in PV systems. For example, a hybrid approach might use unsupervised learning to identify clusters of data and then apply supervised learning to classify the clusters as normal or anomalous.

Regarding the applications of Anomaly Detection in Off-Grid PV Systems the most common use are: [24]

1. Fault Detection: Anomaly detection can be used to identify faults in PV panels, inverters, and other system components. By detecting faults early, PV system operators can take preventive measures to avoid system failures and reduce downtime.

2. Performance Optimization: Anomaly detection can also be used to optimize PV system performance by identifying unusual patterns in data that may indicate suboptimal operating conditions. For example, an anomaly detection algorithm might identify a sudden decrease in energy production and alert the operator to clean the solar panels.

3. Predictive Maintenance: Anomaly detection can be used to predict when maintenance is required, reducing the likelihood of unexpected system failures. By identifying patterns that indicate a failure is imminent, PV system operators can schedule maintenance during off-peak hours, minimizing downtime and improving overall system efficiency.

The challenges of this approach and its future directions can be summarized in: [23]

1. Data Quality: The accuracy of anomaly detection algorithms relies heavily on the quality of the data used for training. However, data quality is often a challenge in off-grid PV systems due to the variability of environmental conditions and the lack of infrastructure for data collection and transmission.

2. False Positives and False Negatives: Anomaly detection algorithms may generate false positives (i.e., identifying a normal pattern as anomalous) or false negatives (i.e., failing to identify an anomaly). Both can have significant consequences in PV systems, and it is crucial to minimize both types of errors.

3. Real-Time Detection: Anomaly detection in off-grid PV systems must be performed in real-time to ensure prompt identification and response to anomalies. This requires advanced computational resources and sophisticated algorithms that can process large amounts of data quickly and accurately.

In conclusion anomaly detection is a critical aspect of off-grid PV system management, enabling the identification of faults, optimization of performance, and predictive maintenance. Various techniques, including supervised and unsupervised learning methods, have been proposed for anomaly detection in PV systems.

While challenges such as data quality and real-time detection remain, advances in machine learning and computational power hold great promise for improving anomaly detection accuracy and efficiency.

3 Methodology

3.1 Description of the study context

The study context of this thesis is focused on the integration of Resilience Engineering Framework (REF) in off-grid renewable energy systems, with a specific focus on photovoltaic (PV) systems. The study aims to investigate the applicability and effectiveness of REF in improving the resilience of off-grid PV systems, and to identify the critical factors that influence the integration of REF in such systems.

The dataset used is an opensource dataset for exploratory data analysis and predictive maintenance studies. Specifically, two different off-grid photovoltaic systems will be studied with the corresponding meteorological data collected by their sensor equipment. This choice makes it possible to go into even more detail on the functioning of the sensor equipment in order to make the two plants independent and to highlight any malfunctions.

The dataset includes information on the power generation data and weather data of the two PV systems collected for a period of 34 days with a 15-minutes sample rate. In particular they are:

- Power generation data:
    - DATE_TIME: Date and time for each observation. Observations recorded at 15 minute intervals.
    - PLANT_ID: this will be common for the entire file
    - SOURCE_KEY: Source key in this file stands for the inverter id.
    - DC_POWER: Amount of DC power generated by the inverter (source_key) in this 15 minute interval. $[kW]$
    - AC_POWER: Amount of AC power generated by the inverter (source_key) in this 15 minute interval. $[kW]$
    - TOTAL_YIELD: This is the total yield for the inverter till that point in time.
- Weather data:
    - DATE_TIME and PLANT_ID already described

- SOURCE_KEY: Stands for the sensor panel id. This will be common for the entire file because there's only one sensor panel for the plant.
- AMBIENT_TEMPERATURE: This is the ambient temperature at the plant [°C]
- MODULE_TEMPERATURE: There's a module (solar panel) attached to the sensor panel. This is the temperature reading for that module. [°C]
- IRRADATION: Amount of irradiation for the 15 minute interval. Note: After comparing this data with other publications, I assume the correct unit for this data is $\left[\frac{kW}{m^2}\right]$

	DATE_TIME	PLANT_ID	SOURCE_KEY	DC_POWER	AC_POWER	DAILY_YIELD	TOTAL_YIELD
0	15-05-2020 00:00	4135001	1BY6WEcLGh8j5v7	0.0	0.0	0.0	6259559.0
1	15-05-2020 00:00	4135001	1IF53ai7Xc0U56Y	0.0	0.0	0.0	6183645.0
2	15-05-2020 00:00	4135001	3PZuoBAID5Wc2HD	0.0	0.0	0.0	6987759.0
3	15-05-2020 00:00	4135001	7JYdWkrLSPkdwr4	0.0	0.0	0.0	7602960.0
4	15-05-2020 00:00	4135001	McdE0feGgRqW7Ca	0.0	0.0	0.0	7158964.0

Fig. 19: Power generation dataset (first 5 rows)

	DATE_TIME	PLANT_ID	SOURCE_KEY	AMBIENT_TEMPERATURE	MODULE_TEMPERATURE	IRRADIATION
0	2020-05-15 00:00:00	4135001	HmiyD2TTLFNqkNe	25.184316	22.857507	0.0
1	2020-05-15 00:15:00	4135001	HmiyD2TTLFNqkNe	25.084589	22.761668	0.0
2	2020-05-15 00:30:00	4135001	HmiyD2TTLFNqkNe	24.935753	22.592306	0.0
3	2020-05-15 00:45:00	4135001	HmiyD2TTLFNqkNe	24.846130	22.360852	0.0
4	2020-05-15 01:00:00	4135001	HmiyD2TTLFNqkNe	24.621525	22.165423	0.0

Fig. 20: Weather dataset (first 5 rows)

Regarding the plant it contains 22 inverters where each inverter is connected with several PV array. Every 15 min, each inverter records his data. So, if we want to know how many the plant has produced a power in an hour, we just compute the contribution of 22 inverters. [24]

To assess the power generated by an off-grid photovoltaic system, monitoring the inverter data is crucial. The inverter is a critical component in the system, as it converts the DC power generated by the solar panels into AC power that can be used by the loads or stored in the battery bank.

The inverter data can provide valuable insights into the system's performance, including the amount of power being generated, the efficiency of the system, and any issues that may arise. By monitoring the inverter data, you can identify patterns and trends that can help you optimize the system's performance and detect potential problems before they become major issues. [25]

To monitor the inverter data, you can use a data logger or a monitoring system that collects data from the inverter and other components of the system, such as the battery bank and load usage.

The data logger or monitoring system can provide real-time data and historical data that can be analyzed to identify trends and patterns in the system's performance.

Some of the key metrics that can be monitored from the inverter data include power output, efficiency, energy yield, voltage, and current. [26] Power output refers to the amount of power being generated by the system at any given time, typically measured in watts (W). Efficiency is the ratio of the amount of power output to the amount of power input, expressed as a percentage. Energy yield refers to the total amount of energy produced by the system over a given period of time, typically measured in kilowatt-hours (kWh). Voltage and current refer to the voltage and current output of the inverter, which can help identify any issues with the system's electrical components. By monitoring these metrics and analyzing the data over time, you can gain valuable insights into the system's performance and identify opportunities for improvement. For example, if you notice that the system's efficiency is lower than expected, you may need to clean the solar panels or adjust the angle of the panels to optimize energy production. [22]

In addition to monitoring the inverter data, it's also important to monitor the system's battery bank and load usage to ensure that the system is operating within safe limits and that the battery bank is being properly maintained. By combining data from the inverter, battery bank, and load usage, you can gain a comprehensive understanding of the system's performance and optimize its operation for maximum efficiency and reliability.

The dataset provides a comprehensive representation of the complex interactions between technical and environmental factors that affect the resilience of off-grid PV systems. By analyzing this dataset through the lens of the Resilience Engineering Framework, this study aims to identify the critical factors that influence the resilience of off-grid PV systems and to develop strategies for improving their resilience."

3.2 Design and Implementation of the Resilience Engineering Framework

In order to establish effective metrics to evaluate the energy resilience of community-level energy systems, this section focuses on performance-based resilience metrics at the system level. A resilience framework is introduced, including a set of metrics that are crucial in the process of decision-making and energy master planning for communities. The present resilience framework utilizes the resilience trapezoid concept [5] to measure (qualitatively and quantitively) the resilience performance of a community-level energy system, proposing a consistent treatment of resilience metrics, dimensions, and phases. This framework is a guide to resilience planning and provides many opportunities for specific resilience enhancement.

These resilience metrics, dimensions, and phases are defined and presented individually. The next section discusses how the presented metrics can work together and be integrated into the resilience planning process during energy master planning to enhance the energy resilience of the community.

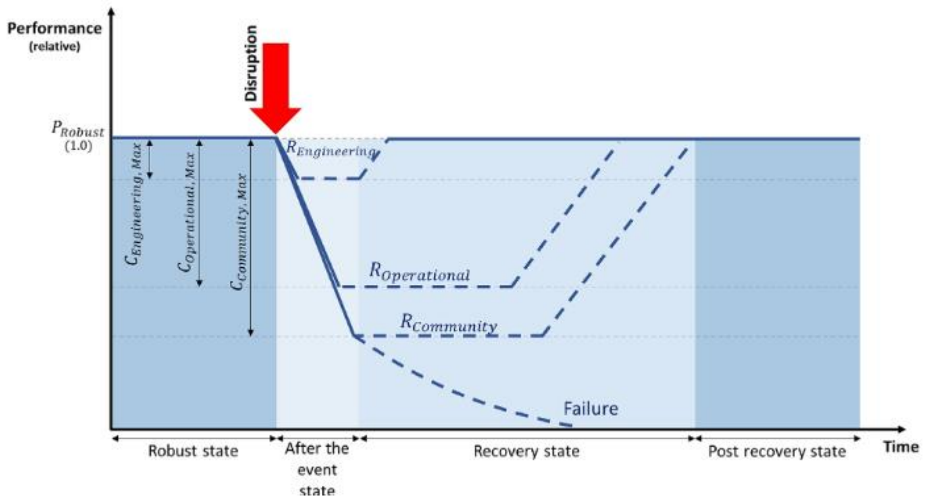

Fig. 21: Conceptual multi-phase energy resilience trapezoid [5]

The multi-phase energy resilience trapezoid is reconceptualized in Fig.21, with different layers. In the robust state, the energy system is working at nearly 100% (or providing reliable energy supply). In the case that an external event causes a noticeable disruption beyond what the system is able to handle, the system's ability to avoid permanent failure and bounce back is illustrated with three layers of resilience [15], working in progression or concurrently, that emphasizes on three dimensions of system resilience performance.

The first layer of resilience is built-in or "engineering-designed resilience" ($R_{Engineering}$). In this approach, the overall energy system assets may be designed in such a way that normal services can be restored after a short disruption, without human intervention. In some sense, this may be seen as an extension of a system designed for reliability and redundancy. But it also provides the opportunity for other or new innovative solutions (e.g. self-healing systems) that are activated when the primary system fails.

The next layer is "operational resilience" ($R\_Operational$), which is the set of technological and organizational measures that can be employed when the disruption exceeds the capacity of engineering-designed resilience.

This also includes the processes of decision-making – from the team or organization level, up to the whole energy sector in a region – that are necessary to contain damage or preserve a certain level of service, and later to fully restore services.

The next level is "community (and societal) resilience" $(R_{Community})$, which needs to be invoked as part of the solution when appropriate, especially when engineering resilience and operational resilience alone are not sufficient to address disruption.

This resilience is defined as "community processes that can restore, maintain or enhance community wellbeing int the face of natural disaster or rapid change".

I extend the definition to include the cooperation, collaboration, or partnership needed between the energy service providers and the demand-side consumers (the "community"). But there is also a broad range of stakeholders within a society, whose cooperation and contributions to manage the disruptions and help bring the service back will be critical. Thus, the community-societal resilience concept can span the range of resilience partners from the direct energy consumers only, to one that includes some or all of the key stakeholders from the wider society interacting with either or both the supply-side industry actors and the demand-side consumers. The three main resilience layers are shown with respect to time in Fig. 21. This period of time is divided into four states including the robust state (before the event), shortly after the event state (including event occurrence state), the recovery state, and the post-recovery state. In each resilience layer, the trapezoid size, slopes, sequence, the length of temporal states and the proportion of presented resilience layers mainly depend on the system performance and type of the disruptive event. Shinozuka, Chang & al. called system resilience capacity (C), "system robustness" which is measured as a percentage. Vugrin and Warren confirm that system resilience capacities can be identified depending on the classes of disruptive events. The summary of layers with examples of the system attributes and enhancement measures are presented in Table 7. In the following sub-sections, these layers are discussed further, and an initial set of metrics are identified for each layer.

	Engineering resilience	Operational resilience	Community resilience
Definition	Physical assets and engineering-designed measures	Set of technological and organisational measures	Cooperation and contributions of customers and other community stakeholders
Example attributes	Redundancy, separation	Rerouting, reorganisation	Awareness, cooperation, trust
Example enhancement measures	Self-healing systems, resistant systems storage systems, energy-self-sufficiency	Demand side management, smart operation, ready supply of critical components	Adaptive thermal comfort, voluntary mass energy-use reduction, mass relocation

Table 7: Summary of energy resilience layers for community-level energy master planning

3.3 Methodologies for integrating anomaly detection into the framework

Monitoring the inverter in a photovoltaic (PV) system is of paramount importance for fault detection and system performance optimization. [27] The inverter plays a critical role in converting solar energy into usable power for the electrical grid, making any malfunction or fault in this component detrimental to the entire system's efficiency and reliability. Through continuous monitoring, potential anomalies or faults can be detected early, allowing for preventive interventions to be implemented before more serious issues arise. This proactive approach not only minimizes downtime but also helps to avoid costly repairs. Additionally, monitoring the inverter enables the identification of operational inefficiencies or malfunctions that could lead to reduced solar energy production. By addressing these issues promptly, optimal system performance can be maintained, ensuring maximum energy output. Moreover, monitoring the inverter is crucial for ensuring the safety of the electrical system, as faults or malfunctions in this component could pose risks such as overvoltages or short circuits, endangering both the system and the individuals involved. The effective monitoring of the inverter is indispensable for ensuring the reliable, efficient, and safe operation of photovoltaic systems.

The PV inverter is the core component of the PV system, and it is essential to develop approaches that accurately predict the occurrence of inverter faults to ensure the PV system’s safety.

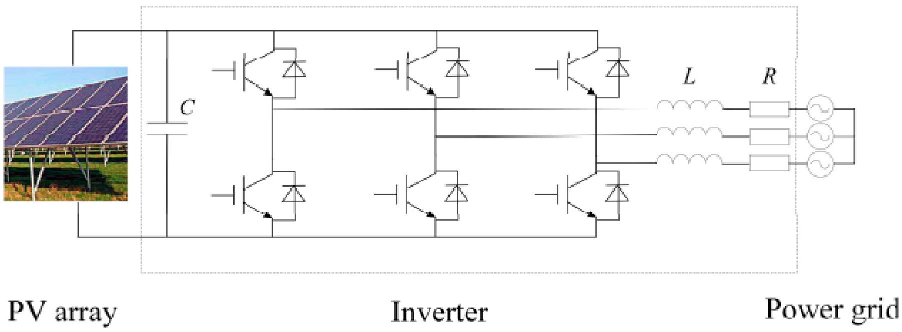

Fig. 22: Main circuit of the photovoltaic (PV) power generation system

In the Table 8 are shown the most common faults events and their cause:

Fault Type	Cause
anomaly of PV string	shielding of PV panel or degradation of PV string
anomaly of DC circuit	DC current protection
anomaly of inverter circuit	inverter current protection
grid connection fault	component damage, etc.
communication fault	communication circuit damaged or disturbed

Table 8: Common faults of the PV power generation system [27]

The methodology developed in this thesis is an application of failure analysis techniques to the individual inverter component instead of the entire photovoltaic system. In particular, as will be seen in detail, we will first count the number of data sockets from the individual inverter (22 will be used) in order to make an initial consideration of the number of data sockets compared to the expected number. The correlation between the various parameters of the dataset will be studied in order to highlight their emerging characteristics. We will go into the details of the evaluations through a study of the residuals of the measured value compared to the predicted one. This will reduce the dimensionality of the problem at hand by studying the single date on which the event occurred. [28]

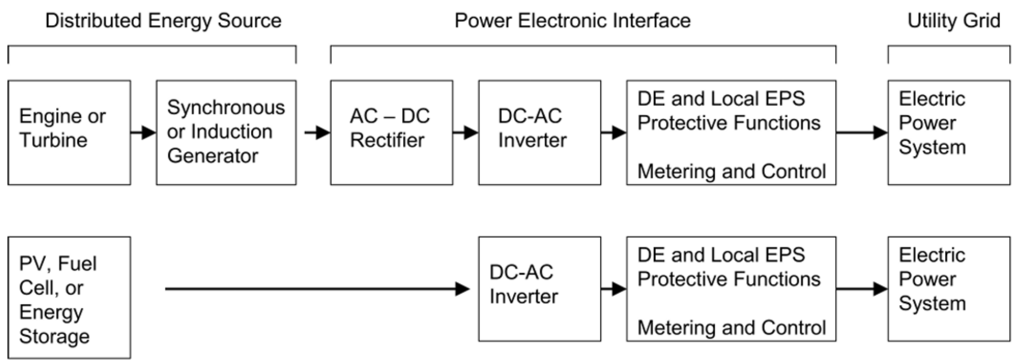

Fig.23: DER system and PE interface block diagram (Kroposki et al. 2006)

4 Design And Implementantion Of Hybrid Ref - Anomaly Detection Algorithm

In this fourth chapter, the type of anomaly detection algorithm used in section 4.1 will be explained.

In section 4.2 below, the correlation with the resilience parameters used is highlighted.

In section 4.3, the dataset used is studied in detail, exposing the characteristics, data collection and initial preparation.

In section 4.4, we expound on how the algorithm was implemented, the development environment and the language used.

In section 4.5, the exploratory data analysis performed and the product of it are reported.

Section 4.6 reports the results and their interpretation in line with what was developed in the previous and following chapters.

4.1 Algorithm Selection for Anomaly Detection

The implementation of the Anomaly Detection algorithm was carried out by means of regression models, linear and non-linear, in order to compare them.

4.1.1 Linear Regression Model

The Linear Regression is a statistical model which estimates the linear relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables). The case of one explanatory variable is called Simple Linear Regression; for more than one, the process is called Multiple Linear Regression. This term is distinct from multivariate linear regression, where multiple correlated dependent variables are predicted, rather than a single scalar variable. If the explanatory variables are measured with error then errors-in-variables models are required, also known as measurement error models. [29]

In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models. Most commonly, the conditional mean of the response given the values of the explanatory variables (or predictors) is assumed to be an affine function of those values; less commonly, the conditional median or some other quantile is used.

Like all forms of regression analysis, linear regression focuses on the conditional probability distribution of the response given the values of the predictors, rather than on the joint probability distribution of all of these variables, which is the domain of multivariate analysis.

Linear regression has many practical uses. Most applications fall into one of the following two broad categories:

- If the goal is error i.e variance reduction in prediction or forecasting, linear regression can be used to fit a predictive model to an observed data set of values of the response and explanatory variables. After developing such a model, if additional values of the explanatory variables are collected without an accompanying response value, the fitted model can be used to make a prediction of the response.
- If the goal is to explain variation in the response variable that can be attributed to variation in the explanatory variables, linear regression analysis can be applied to quantify the strength of the relationship between the response and the explanatory variables, and in particular to determine whether some explanatory variables may have no linear relationship with the response at all, or to identify which subsets of explanatory variables may contain redundant information about the response.

Linear regression models are often fitted using the least squares approach, but they may also be fitted in other ways, such as by minimizing the "lack of fit" in some other norm (as with least absolute deviations regression), or by minimizing a penalized version of the least squares cost function as in ridge regression (L2-norm penalty) and lasso (L1-norm penalty). Use of the Mean Squared Error (MSE) as the cost on a dataset that has many large outliers, can result in a model that fits the outliers more than the true data due to the higher importance assigned by MSE to large errors. So, cost functions that are robust to outliers should be used if the dataset has many large outliers. Conversely, the least squares approach can be used to fit models that are not linear models. Thus, although the terms "least squares" and "linear model" are closely linked, they are not synonymous.

Given a dataset $\{y_i, x_{i1}, \dots, x_{ip}\}_{i=1,\dots,n}$ of n statistical units, a linear regression model assumes that the relationship between the dependent variable y and the vector of regressor $\boldsymbol{x}$ is linear. This relationship is modeled through a disturbance term or error variable ϵ – an unobserved random variable that adds "noise" to the linear relationship between the dependent variable and regressors. Thus the model takes the form:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \epsilon_i \qquad (5)$$

With $i = 1, .., n$, where α^T denote the transpose, so that $\boldsymbol{x}_i^T \boldsymbol{\beta}$ is the inner product between vectors $\boldsymbol{x}_i$ and $\boldsymbol{\beta}$.

Often these n equations are stacked together and written in matrix notation:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon} \qquad (6)$$

Where:

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ .. \\ y_n \end{bmatrix}, \boldsymbol{X} = \begin{bmatrix} x_1^T \\ .. \\ x_n^T \end{bmatrix} = \begin{bmatrix} 1 & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ 1 & \cdots & x_{np} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ .. \\ \beta_p \end{bmatrix}, \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ .. \\ \epsilon_n \end{bmatrix}$$

- y_i is a vector of observed values ot the variable called the regressand or measured variable/dependent variable. This variable is also sometimes known as the predicted variable, but this should not be confused with predicted values, which are denoted $\hat{y}$. The decision as to which variable in a dataset is modeled as the dependent variable and which are modeled as the independent variable may be based on a presumption that the value of one of the variable is caused by, or directly influenced by the other variables. Alternatively, there may be an operational reason to model one of the variables in terms of the others, in which case there need be no presumption of causality.
- $\boldsymbol{X}$ be seen as a matrix of row-vectors $\boldsymbol{x_{i\cdot}}$ or of n-dimensional column-vectors $\boldsymbol{x_{\cdot j}}$, which are known as regressors, predictor variables or independent variables (not to be confused with the concept of independent random variables). The matrix $\boldsymbol{X}$ is sometimes called the design matrix.
  - Usually, a constant is included as one of the regressors. In particular, $x_{i0} = 1$ for $i = 1, .., n$. The corresponding element of $\boldsymbol{\beta}$ is called the intercept. Many statistical inference procedures for linear models require an intercept to be present, so it is often included even if theoretical considerations suggest that its value should be zero.
  - Sometimes one of the regressors can be a non-linear function or another regressor or of the data values, as in polynomial regression and segmented regression. The model remains linear as long as it is linear in the parameter vector $\boldsymbol{\beta}$.
  - The values x_{ij} may be viewed as either observed values of random variables X_j or as fixed values chosen prior to observing the dependent variable. Both interpretations may be appropriate in different cases, and they generally lead to the same estimation procedures; however different approaches to asymptotic analysis are used in these two situations.
- $\boldsymbol{\beta}$ is a $(p + 1)$ – dimensional parameter vector, where β_0 is the intercept term (if one is included in the model, otherwise is p-dimensional). Its elements are known as effects of regression coefficients (although the latter term is sometimes reserved for the estimated effects). In simple linear regression, $p = 1$, and the coefficient I known as regression slope. Statistical estimation and inference in linear regression focuses

on $\boldsymbol{\beta}$. The elements of this parameter vector are interpreted as the partial derivatives of the dependent variable with respect to the various independent variables.

- $\boldsymbol{\epsilon}$ is a vector of values ϵ_i. This part of the model is called error term or disturbance term, or sometimes noise (in contrast with the "signal" provided by the rest of the model). This variable captures all other factors which influence the dependent variable y other than the regressors $\boldsymbol{x}$. The relationship between the error term and the regressors, for example their correlation, is a crucial consideration in formulating a linear regression model, as it will determine the appropriate estimation method.

Fitting a linear model to a given dataset usually requires estimating the regression coefficients $\boldsymbol{\beta}$ such that the error term $\boldsymbol{\epsilon} = \boldsymbol{y} - \boldsymbol{X\beta}$ is minimized. For example, it is common to use the sum of squared errors $||\epsilon||_2^2$ as a measure of $\boldsymbol{\epsilon}$ for minimization.

A large number of procedures have been developed for parameter estimation and inference in linear regression. These methods differ in computational simplicity of algorithms, presence of a closed-form solution, robustness with respect to heavy-tailed distributions, and theoretical assumptions needed to validate desirable statistical properties such as consistency and asymptotic efficiency.

The technique that has been used is Least-Squares Estimation, exposed in the following paragraph.

Assuming that the independent variable is $\boldsymbol{x_i} = \left[x_1^i, x_2^i, \ldots, x_m^i\right]$ and the model's parameters are $\boldsymbol{\beta} = [\beta_0, \beta_1, \ldots, \beta_m]$ then the model's prediction would be:

$$y_i \approx \beta_0 + \sum_{j=1}^{m} \beta_j x_j^i \qquad (7)$$

If $\boldsymbol{x_i}$ is extended to $\boldsymbol{x_i} = \left[1, x_1^i, \ldots, x_m^i\right]$ then y_i would become a dot product of the parameter and the independent variable, i.e.:

$$y_i \approx \sum_{j=0}^{m} \beta_j x_j^i \; = \boldsymbol{\beta x_i} \qquad (8)$$

In the least-squares setting, the optimum parameter is defined as such that minimizes the sum of mean squared loss:

$$\widehat{\boldsymbol{\beta}} = argmin_{\beta} L(D, \boldsymbol{\beta}) = argmin_{\beta} \sum_{i=1}^{n} (\boldsymbol{\beta}\boldsymbol{x_i} - y_i)^2 \quad (9)$$

Now putting the independent and variables in matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ respectively, the loss function can be rewritten as:

$$L(D, \boldsymbol{\beta}) = \|\boldsymbol{X\beta} - \boldsymbol{Y}\|^{\boldsymbol{2}}$$

$$= (\boldsymbol{X\beta} - \boldsymbol{Y})^T(\boldsymbol{X\beta} - \boldsymbol{Y})$$

$$= \boldsymbol{Y^T Y} - \boldsymbol{Y^T X\beta} - \boldsymbol{\beta^T X^T Y} + \boldsymbol{\beta^T X^T X\beta} \qquad (10)$$

As the loss is convex the optimum solution lies at gradient zero. The gradient of the loss function is (using Denominator layout convention):

$$\frac{\partial L(D, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{\partial(\boldsymbol{Y^T Y} - \boldsymbol{Y^T X\beta} - \boldsymbol{\beta^T X^T Y} + \boldsymbol{\beta^T X^T X\beta})}{\partial \boldsymbol{\beta}}$$

$$= -2\boldsymbol{X^T Y} + 2\boldsymbol{X^T X\beta} \qquad (11)$$

Setting the gradient to zero produces the optimum parameter:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X^T X})^{-\boldsymbol{1}}\boldsymbol{X^T Y} \qquad (12)$$

Note that to prove that the $\widehat{\boldsymbol{\beta}}$ obtained is indeed the local minimum, one needs to differentiate once more to obtain the Hessian matrix and show that it is positive definite. This is provided by the Gauss-Markov theorem. [30]

4.1.2 Non-Linear Regression Model

The nonlinear regression is a form of regression analysis in which observational data are modeled by a function which is a nonlinear combination of the model parameters and depends on one or more independent variables. The data are fitted by a method of successive approximations (iteration). [29]

In nonlinear regression a statistical model of the form:

$$\boldsymbol{y} \sim f(\boldsymbol{x}, \boldsymbol{\beta}) \quad (13)$$

Relates a vector of independent variables, $\boldsymbol{x}$**,** and its associated observed dependent variables, $\boldsymbol{y}$**.** The function f is non linear in the components of the vector of parameters β_i, but otherwise arbitrary.

In general, there is no closed-form expression for the best-fitting parameters, as there is in linear regression. Usually, numerical optimization algorithms are applied to determine the best-fitting parameters. Again, in contrast to linear regression, there may be many local minima of the function to be optimized and even the global minimum may produce a biased estimate. In practice, estimated values of the parameters are used, in conjunction with the optimization algorithm, to attempt to find the global minimum of a sum of squares.

The assumption underlying this procedure is that the model can be approximated by a linear function, namely first-order Taylor series:

$$f(x_i, \beta) \approx f(x_i, 0) + \sum_j J_{ij}\beta_j \qquad (14)$$

Where the term $J_{ij} = \frac{\partial f(x_i,\beta)}{\partial \beta_j}$ are Jacobian matrix elements. It follows from this that the least squares estimators are given by:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{J}^{\boldsymbol{T}}\boldsymbol{J})^{-\boldsymbol{1}}\boldsymbol{J}^{\boldsymbol{T}}\boldsymbol{y} \qquad (15)$$

Compare generalized least squares with covariance matrix proportional to the unit matrix. The nonlinear regression statistics are computed and used as in linear regression statistics but using **J** in place of **X** in the formulas.

When the function (14) itself is not known analytically, but needs to be linearly approximated from $n + 1$, or more, known values (where n is the number of estimators), the best estimator is obtained directly from the Linear Template Fit as:

$$\widehat{\boldsymbol{\beta}} = \left(\left(\boldsymbol{Y}\widetilde{\boldsymbol{M}}\right)^{\boldsymbol{T}}\boldsymbol{\Omega}^{-\boldsymbol{1}}\boldsymbol{Y}\widetilde{\boldsymbol{M}}\right)^{-\boldsymbol{1}}\left(\boldsymbol{Y}\widetilde{\boldsymbol{M}}\right)^{\boldsymbol{T}}\boldsymbol{\Omega}^{-\boldsymbol{1}}(\boldsymbol{d} - \boldsymbol{Y}\overline{\boldsymbol{m}}) \qquad (16)$$

The linear approximation introduces bias into the statistics. Therefore, more caution than usual is required in interpreting statistics derived from a nonlinear model. The best-fit curve is often assumed to be that which minimizes the sum of squared residuals. This is the ordinary least squares (OLS) approach. However, in cases where the dependent variable does not have constant variance, or there are some outliers, a sum of weighted squared residuals may be minimized. Each weight should ideally be equal to the reciprocal of the variance of the observation, or the reciprocal of the dependent variable to some power in the outlier case, but weights may be recomputed on each iteration, in an iteratively weighted least squares algorithm. [29]

4.1.3 Regression Techniques in Anomaly Detection

In this study, I employed both linear and nonlinear regression techniques to detect anomalies in the dataset. Linear regression was used to identify patterns in the data that followed a linear trend, while nonlinear regression was used to identify patterns that deviated from a linear trend. By combining both linear and nonlinear regression techniques, I was able to detect a wider range of anomalies that may have been missed by using a single approach. The combination of both models provided a more comprehensive understanding of the data and improved the accuracy of anomaly detection.

The combination of linear and nonlinear regression techniques provides a powerful tool for anomaly detection in various fields. The hybrid approach can identify both linear and nonlinear patterns in the data, which can improve the accuracy of anomaly detection. This approach can be used in various applications where accurate anomaly detection is critical, such as fraud detection, network intrusion detection, and quality control.

In particular the linear regression can be impacted by outliers in two ways:

1. An extreme outlier in the y-direction at x-values near $\bar{x}$ can affect the fit in that area in the same way an outlier can affect a mean
2. An "outlying" observation in x-space is an influential observation – it can pull the fit of the line toward it. If it is sufficiently far away the line will go through the influential point:
3.

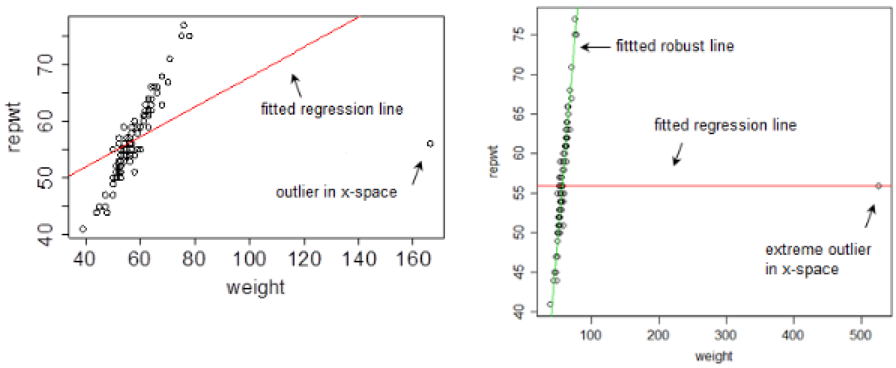

Fig. 24: Implementation of linear regression for outliers' detection [29]

In the left plot, there's a point that's quite influential, and it pulls the line quite a way from the large bulk of the data. In the right plot, it's been moved even further away -- and now the line goes through the point. When the x-value is that extreme, as you move that point up and down, the line moves with it, going through the mean of the other points and through the one influential point.

An influential point that's perfectly consistent with the rest of the data may not be such a big problem, but one that's far from a line through the rest of the data will make the line fit it, rather than the data.

Looking at the right-hand plot, the red line - the least squares regression line - doesn't show the extreme point as an outlier at all - its residual is 0. Instead, the large residuals from the least squares line are in the main part of the data! This means you can completely miss an outlier.

Even worse, with multiple regression, an outlier in x-space may not look particularly unusual for any single x-variable. If there's a possibility of such a point, it's potentially a very risky thing to use least squares regression on.

4.2 Resilience Indicators

4.2.1 Engineering-designed resilience metrics

Engineering-designed resilience enhancement strategies at the asset level, are usually of physical/protective/mechanical nature, and do not require human intervention to apply. [16]. Some example measures are redundant capacities, storage, backup systems, and physicial protective measures. Several metrics for engineering-designed resilience (also called infrastructure/asset resilience in some literature) have been proposed in Fig. 21, the slope and extent of the edges of engineering resilience trapezoid (three pieces) and integral under the curve (area) are example metrics. The maximum engineering-designed resilience capacity ($C_{Engineering,Max}$) of the system represents the overall resilience capacity of the assets. Examples of engineering-designed resilience metrics for power networks include:

- the rate of disturbance (number of lines tripped per hour and number of lines tripped)

- the duration of the performance distruption (hours)

- the rate of system recovery (number of lines restored)

Other examples of engineering resilience metrics that can be used in the energy master planning and design of communities are energy (kWh) not served due to assets failure, and asset availability (measured by the amount of time the asset serves its intended purpose divided by the total amount of time the asset was exposed to disruptive conditions). These metrics can be used to estimate the energy resilience of a community system against events with similar probability and intensity.

4.2.2 Operational Resilience Metrics

Operational resilience is focused on system level performance an the operational characteristics of the system intended to mitigate the failure risk and to support service recovery. Some examples of enhancement measures are demand response, demand side management (DSM) strategies, prioritizing energy use, smart controls and forecasting. The slope and extent of the edges of the operational resilience trapezoid and integral under the curve are examples of metrics that can be used to measure the operational resilience of the system. The maximum capacity of operational resilience ($C_{Operational,Max}$) represents the maximum level of energy resilience that the system can achieve through operational measures. For example, in power networks operational resilience metrics include:

- The rate of disturbance (kW power loss per hour and power capacity loss)
- The duration of the performance disruption (hours)
- The rate of system recovery (kW power restored)

Other examples of operational resilience metrics are energy (kWh) not served due to operation disruption and energy availability (measured by the amount of energy served to end users divided by the total amount energy demand by those users during disruptive conditions).

4.2.3 Community Resilience Metrics

Community-societal energy resilience accounts for the actions that should be done within the community by some or all of community users to maintain the minimum allowable community-societal services. Mass relocation, effective use of community resources during a disruption, and increasing the bonding, bridging, and linking the social capital, are examples of community-societal resilience enhancement measures.

The type of these civil actions can be an adaption or halting of normal actions and can vary depending on the disruptive event and the community wishes. These are especially critical when the minimum engineering-designed and operational resilience limits have failed. Compared to the previous two layers, community-societal resilience usually has a much higher resilience capacity ($C_{Community,Max}$). The lack of critical services impacts both the occupants of the community and also the greater society. Here, the focus is on both of these impacts.

The maximum capacity of community-societal resilience of the system is very challenging if not impossible to ascertain. An alternative is to use a community resilience metrics such as the community functionality which is measured by the amount of time-critical community functions (e.g. energy services) were adequately provided to people, divided by the total amount of disruption time. This could be separated out by function or combined in a weighted manner.

4.3 Data Collection and Preparation

As shown in the paragraph 3.3 the dataset includes information on the power generation data and weather data of the two PV systems collected for a period of 34 days with a 15-minutes sample rate.

In particular they are:

- Power generation data:
    - DATE_TIME: Date and time for each observation. Observations recorded at 15 minute intervals.
    - PLANT_ID: this will be common for the entire file
    - SOURCE_KEY: Source key in this file stands for the inverter id.
    - DC_POWER: Amount of DC power generated by the inverter (source_key) in this 15 minute interval. $[kW]$
    - AC_POWER: Amount of AC power generated by the inverter (source_key) in this 15 minute interval. $[kW]$
    - TOTAL_YIELD: This is the total yield for the inverter till that point in time.
- Weather data:
    - DATE_TIME and PLANT_ID already described

- SOURCE_KEY: Stands for the sensor panel id. This will be common for the entire file because there's only one sensor panel for the plant.
- AMBIENT_TEMPERATURE: This is the ambient temperature at the plant [°C]
- MODULE_TEMPERATURE: There's a module (solar panel) attached to the sensor panel. This is the temperature reading for that module. [°C]
- IRRADATION: Amount of irradiation for the 15 minute interval. Note: After comparing this data with other publications, I assume the correct unit for this data is $\left[\frac{kW}{m^2}\right]$

In the realm of renewable energy, photovoltaic (PV) systems play a crucial role in harnessing solar energy for electricity generation. Monitoring and analyzing the performance of these systems are essential for optimal operation, maintenance, and efficiency improvement. Data acquisition from PV systems involves the collection, processing, and analysis of various parameters to ensure their effective functioning and performance optimization.

Data Acquisition Components:

1. Sensors: Sensors are deployed within the PV system to measure key parameters such as solar irradiance, ambient temperature, panel temperature, voltage, current, and power output. These sensors can be integrated into the PV panels, inverters, or placed strategically within the system.

2. Data Loggers: Data loggers are electronic devices responsible for recording data from the sensors at predefined intervals. They typically feature multiple channels to accommodate various sensor inputs and store data in digital format for further analysis.

3. Communication Interfaces: Communication interfaces enable the transfer of data from data loggers to a centralized monitoring system or cloud-based platform. Common communication protocols include Modbus, TCP/IP, and MQTT, facilitating seamless integration with monitoring software.

Data Acquisition Process:

1. Sensor Calibration: Before deployment, sensors undergo calibration to ensure accurate and reliable measurements. Calibration involves adjusting sensor outputs to match known reference values under specific conditions, thereby minimizing measurement errors.

2. Data Logging: Data loggers record sensor readings at regular intervals, typically ranging from minutes to hours, depending on the monitoring requirements. Recorded data may include irradiance levels, temperature variations, voltage, current, and power output.

3. Data Transmission: Once logged, data is transmitted from the data loggers to a central repository or monitoring platform. This transmission can occur via wired (Ethernet, RS-485) or wireless (Wi-Fi, cellular) communication channels, enabling real-time monitoring and analysis.

4. Data Analysis: In the monitoring platform, collected data undergoes analysis to assess the performance and health of the PV system. Key performance indicators (KPIs) such as energy yield, efficiency, and degradation rate are calculated and compared against expected values or historical data.

5. Reporting and Visualization: Analysis results are presented through intuitive dashboards, graphs, and reports, providing insights into system performance trends, anomalies, and potential issues. Visualization tools enhance data interpretation and facilitate informed decision-making regarding maintenance, troubleshooting, and performance optimization.

The major benefits of Data Acquisition are:

1. Performance Optimization: Continuous monitoring and analysis enable proactive identification of performance issues, allowing for timely maintenance and optimization measures to maximize energy yield and efficiency.

2. Fault Detection and Diagnosis: Data acquisition facilitates early detection of system faults or anomalies, enabling prompt troubleshooting and resolution to minimize downtime and revenue losses.

3. Asset Management: Comprehensive data on PV system performance and health support effective asset management strategies, including lifecycle planning, warranty management, and investment decision-making.

4. Compliance and Reporting: Accurate data acquisition ensures compliance with regulatory requirements and standards, facilitating reporting obligations and audits for stakeholders such as utilities, regulators, and investors.

Effective data acquisition is integral to maximizing the performance, reliability, and longevity of photovoltaic (PV) systems. By leveraging advanced sensors, data loggers, and communication technologies, stakeholders can gather actionable insights to optimize energy production, reduce operational costs, and contribute to a sustainable energy future.

Having explained the data acquisition part in general, we go into the details of the pre-processing performed.

A first check on the dataset is to write a part of the algorithm that iteratively checks each element and its corresponding PLANT_ID in order to verify the uniqueness of the belonging plant. The result is a set of {1} symbols that the plant value exists and is unique.

A first check on the cleanliness of the dataset involves checking for missing values in the dataset. This is a condition to be remedied either by deleting the row/column corresponding to the missing data or by replacing it with statistically relevant values such as mean or median.

```
DATE_TIME      0
PLANT_ID       0
SOURCE_KEY     0
DC_POWER       0
AC_POWER       0
DAILY_YIELD    0
TOTAL_YIELD    0
dtype: int64
```

```
DATE_TIME              0
PLANT_ID               0
SOURCE_KEY             0
AMBIENT_TEMPERATURE    0
MODULE_TEMPERATURE     0
IRRADIATION            0
dtype: int64
```

Fig.25: Output of data cleaning code regarding missing values

The strategy adopted to check for missing values was to sum up all elements or within each individual column of the dataset. As can be seen in Fig.25, the check revealed their absence and thus a degree of cleanliness in the dataset.

A closer check is made on the individual inverters and the number of acquisitions is evaluated. The 22 inverters are coded in order to have a unique correspondence. The sum of the number of acquisitions is carried out and this defines in the first instance a possible failure information of the inverters. Should the number of acquisitions be lower than expected, this could be a symptom of a malfunction.

```
SOURCE_KEY
bvBOhCH3iADSZry    3155
1BY6WEcLGh8j5v7    3154
7JYdWkrLSPkdwr4    3133
VHMLBKoKgIrUVDU    3133
ZnxXDlPa8U1GXgE    3130
ih0vzX44oOqAx2f    3130
z9Y9gH1T5YWrNuG    3126
wCURE6d3bPkepu2    3126
uHbuxQJl8lW7ozc    3125
pkci93gMrogZuBj    3125
iCRJl6heRkivqQ3    3125
rGa61gmuvPhdLxV    3124
sjndEbLyjtCKgGv    3124
McdE0feGgRqW7Ca    3124
zVJPv84UY57bAof    3124
ZoEaEvLYb1n2sOq    3123
1IF53ai7Xc0U56Y    3119
adLQvlD726eNBSB    3119
zBIq5rxdHJRwDNY    3119
WRmjgnKYAwPKWDb    3118
3PZuoBAID5Wc2HD    3118
YxYtjZvoooNbGkE    3104
Name: count, dtype: int64
```

Fig.26: number of acquisitions per inverter (SOURCE_KEY)

As observed, there are 22 distinct inverters with measurement counts ranging from 3104 to 3155. This variance could potentially pose challenges for prediction models and needs to be considered. Given that each entry represents a 15-minute measurement interval, the maximum deviation of 51 entries translates to a discrepancy of nearly 13 hours.

4.4 Exploratory Data Analysis

The correlation matrix, denoted as $\boldsymbol{R}$, is a fundamental tool in statistics and data analysis, particularly in the field of multivariate analysis. [29] It succinctly summarizes the pairwise correlations between variables in a dataset. Mathematically, the correlation between variable x_i and x_j can be expressed as:

$$\rho_{ij} = \frac{Cov(x_i, x_j)}{\sigma_i \sigma_j} \quad (17)$$

Where ρ_{ij} represents the correlation coefficient between x_i and x_j, $Cov(x_i, x_j)$ denotes the covariance between x_i and x_j, and σ_i and σ_j are the standard deviations of x_i and x_j respectively.

In particular the standard deviation is defined by:

$$\sigma_i = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \qquad (18)$$

Where:

- N is the total number of observations

- x_i represents each individual observation of the variable

- μ is the mean of the variable

While the $Cov(x_i, x_j)$ is expressed by:

$$Cov(x_i, x_j) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_{x_i})(x_j - \mu_{x_j}) \qquad (19)$$

Where:

- N is the total number of observations

- x_i and x_j represents each individual observation

- μ is the mean of the variable

The correlation matrix $\boldsymbol{R}$ itself is a symmetric matrix with diagonal elements equal to 1, as the correlation of a variable itself is always 1. It provides valuable insights into the relationships between variables: positive values indicate a positive linear relationship, negative values indicate a negative linear relationship, and values close to zero suggest a weak or no linear relationship. Additionally, the correlation matrix aids in identifying multicollinearity, where two or more variables are highly correlated, potentially leading to issues in regression analysis such as inflated standard errors and unreliable coefficient estimates. By examining the correlation matrix, we can pinpoint highly correlated variables and make informed decisions regarding variable selection or model specification. In Figure 27 is shown the $\boldsymbol{R}$ between the jointed values of power and weather dataset for each plant.

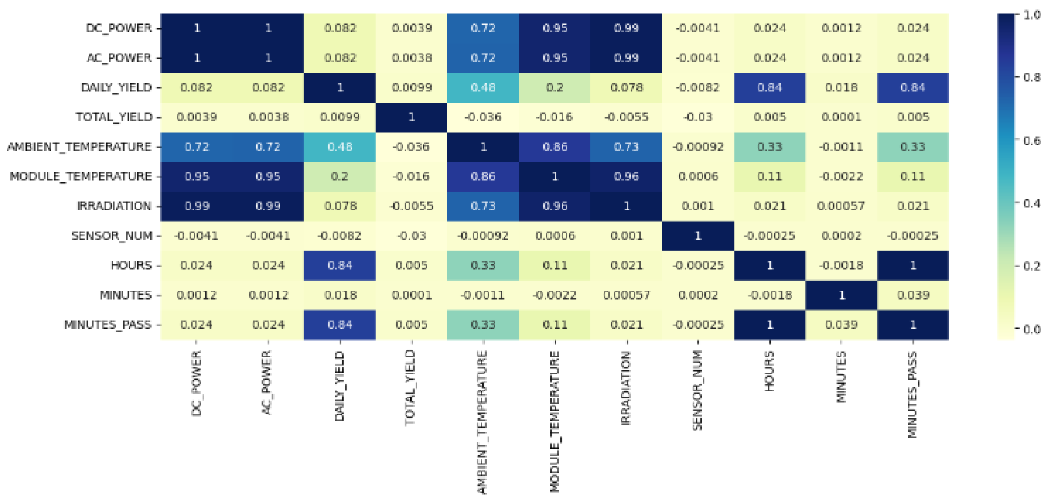

Fig. 27: Correlation matrix model applied to power and weather dataset

From the correlation matrix is possible to gain a lot of insight:

- high correlation between DC Power and AC Power
- high correlation between Power and Irradiation
- correlation between DC Power, AC Power and Module Temperature and Ambient Temperature
- correlation between Daily Yield and Ambient Temperature

In Figure 28 is exposed a more comprehensive look at the correlation between AC and DC power with respect to themselves and the Irradiation. In particular the list of inverter is reported in the right-hand side in order to underline which inverter is candidate for a more complete investigation to find faults.

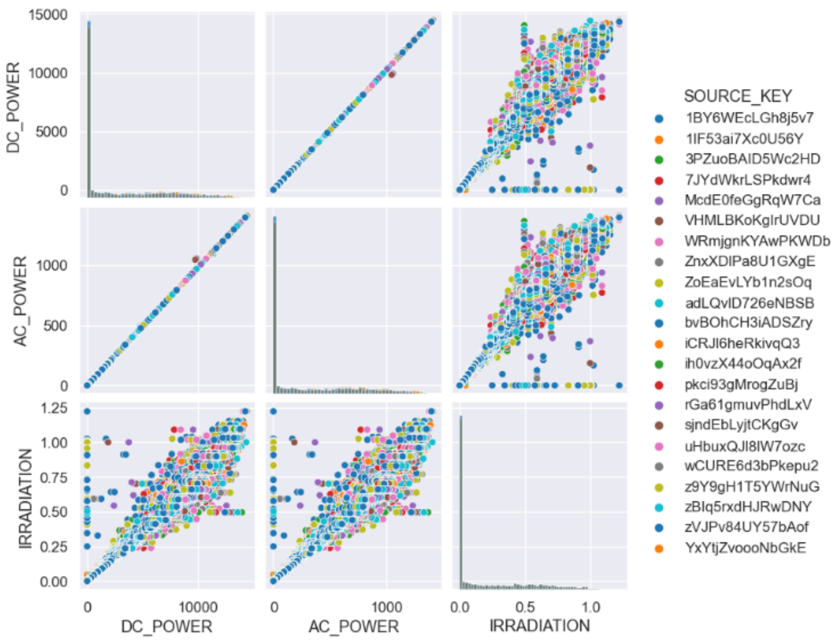

Fig.28: Correlation matrix between AC, DC Power and Irradiation

The correlation analysis between DC Power and AC Power reveals a strong correlation close to unity, indicating that they typically vary together. However, there are slight deviations observed in some inverters, suggesting potential anomalies or irregularities in their performance.

In contrast, when examining the relationship between Irradiation and DC Power, while there is still a positive correlation, it is noticeably weaker compared to the correlation between DC Power and AC Power. Despite this positive correlation, the spread of data points is wider, indicating greater variability in DC Power with changes in Irradiation.

Interestingly, it's worth noting that instances occur where there is irradiation present, but the inverters register zero power output, whether in DC or AC. This discrepancy, where there is energy available but not being converted, points towards potential malfunctioning inverters or issues within the system that need further investigation.

Outliers in Power-Irradiation indicate failure of the panel lines. If there is enough sunlight but no power is generated, this points to faulty photovoltaic cells.

Outliers in DC-AC conversion indicate failure at the inverter. If there is DC power delivered but less AC power generated than expected the inverter may be malfunctioning calculated from measured DC_POWER, there seems to be an issue with how this data was generated.

Let's look a bit closer at the pairplots where we identified outliers and see if these are spread out evenly across all inverters or if we can identify specific inverters.

In Figure 29, the relationship between AC Power and DC Power is depicted, showcasing the efficiency of our inverter in converting DC power to AC power. This graphical representation provides valuable insights into the performance of the conversion process. Anomalies or irregularities observed in this relationship can serve as significant indicators for detecting malfunctioning inverters within the system. By closely examining fluctuations or discrepancies between AC and DC power outputs, potential issues can be identified and addressed promptly, ensuring the smooth operation and optimal performance of the entire power conversion system.

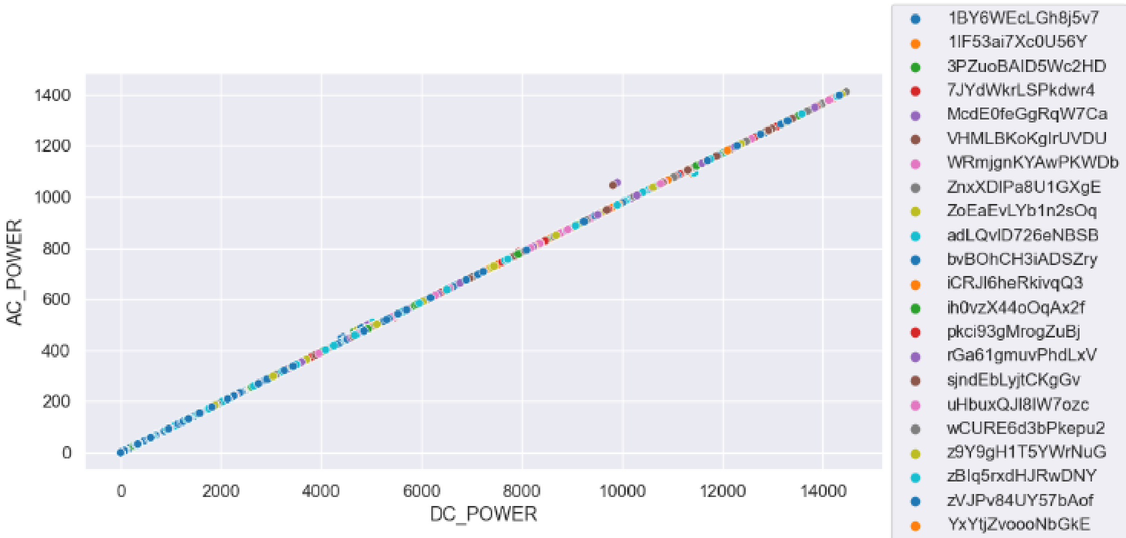

Fig. 29: AC Power and DC Power correlation

In Figure 30 a detailed examination reveals the presence of outliers particularly noticeable in the relationship between DC Power and AC Power concerning Module Temperature. While other temperature variables don't exhibit any glaring outliers, the DC&AC Power versus Module Temperature plot distinctly highlights data points that deviate significantly from the overall pattern. These outliers in power generation concerning module temperature could signify instances of equipment malfunction, environmental factors affecting performance, or other anomalies warranting closer investigation.

Identifying and understanding these outliers is crucial for maintaining system efficiency and reliability, as well as ensuring accurate data analysis and interpretation in the context of power generation systems.

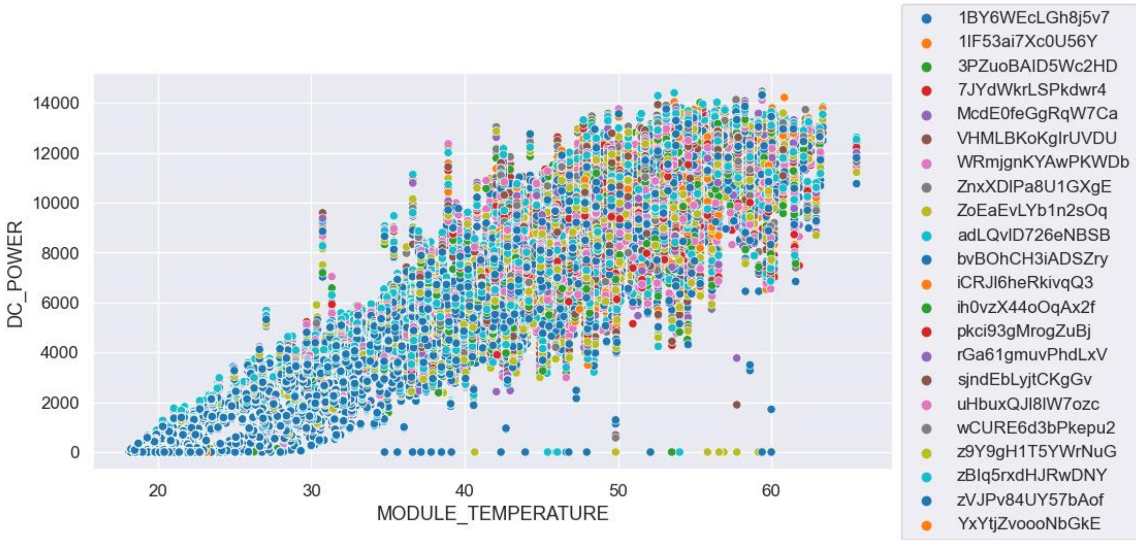

Fig. 30: DC Power and Module Temperature

The observation in Figure 31 underscores the effectiveness of our photovoltaic panel lines in efficiently transforming sunlight into DC power. The outliers detected during this process of converting solar energy into electricity are indicative of potential issues with malfunctioning photovoltaic panel lines. These anomalies may stem from various factors such as panel degradation, shading, soiling, or even manufacturing defects. Identifying and addressing these anomalies promptly is crucial to maintain optimal performance and maximize energy output from the photovoltaic system. By closely monitoring the relationship between sunlight intensity, module temperature, and DC power output, we can pinpoint areas of concern and take proactive measures to rectify any issues, ensuring the long-term reliability and sustainability of our solar energy infrastructure.

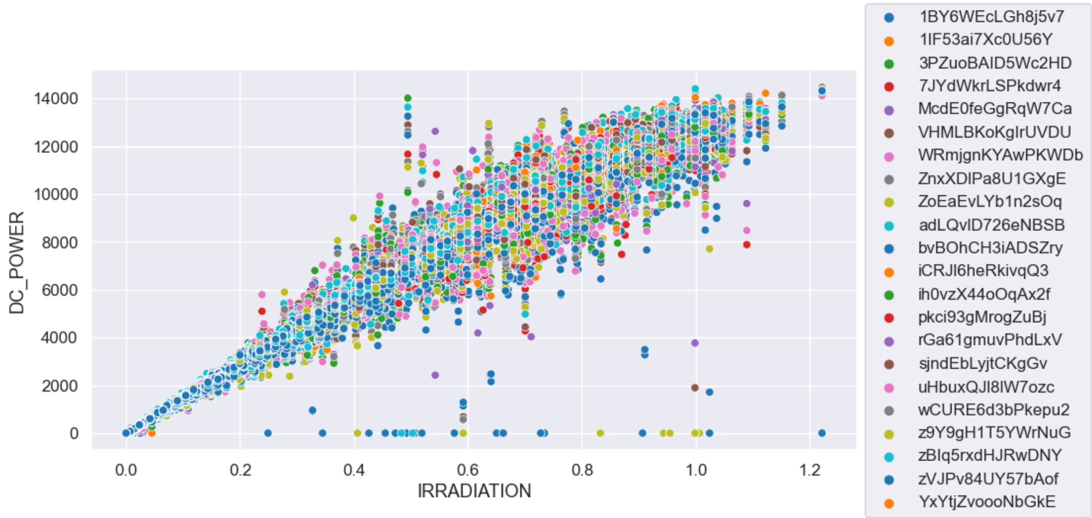

Fig. 31: DC Power and Irradiation correlation

Our dataset vividly illustrates occurrences where certain inverters received no DC power despite ample sunlight available for power generation. These instances unequivocally indicate equipment malfunction within our system. To provide a visual representation of this issue, we can delve deeper into the daily distribution of generated power and the measured irradiation.

By examining the daily distribution of generated power alongside the measured irradiation levels, we can uncover discrepancies and anomalies that point towards malfunctioning equipment. This comparative analysis will help us identify specific time periods or conditions where inverters fail to convert available sunlight into electrical power, shedding light on the extent and frequency of the malfunctioning events.

Furthermore, visualizing this data will facilitate a comprehensive understanding of the relationship between irradiation levels and power generation, enabling us to pinpoint and address the root causes of the equipment malfunction effectively.

This proactive approach is essential for maintaining the efficiency, reliability, and performance of our solar energy infrastructure, ultimately ensuring uninterrupted power generation and maximizing our renewable energy output.

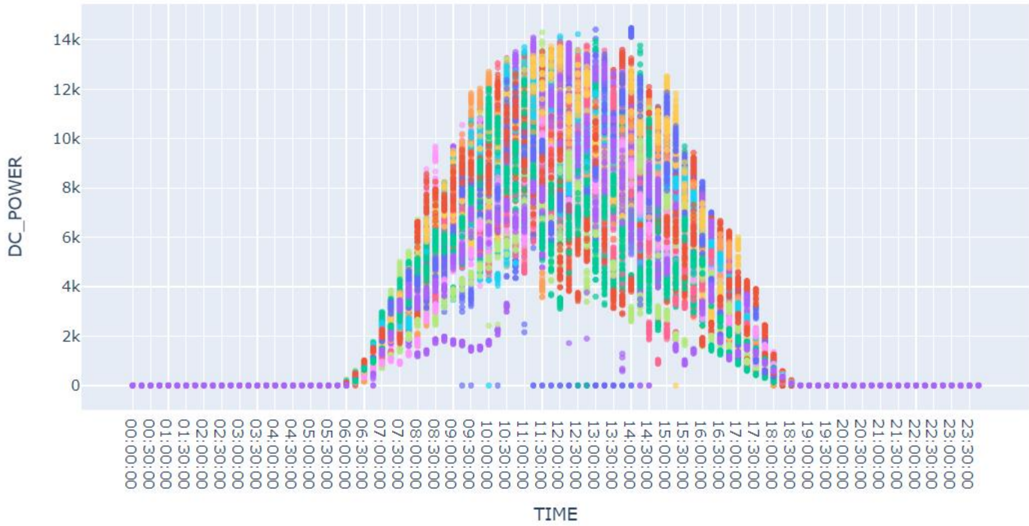

Fig. 32: Average DC Power daily distribution

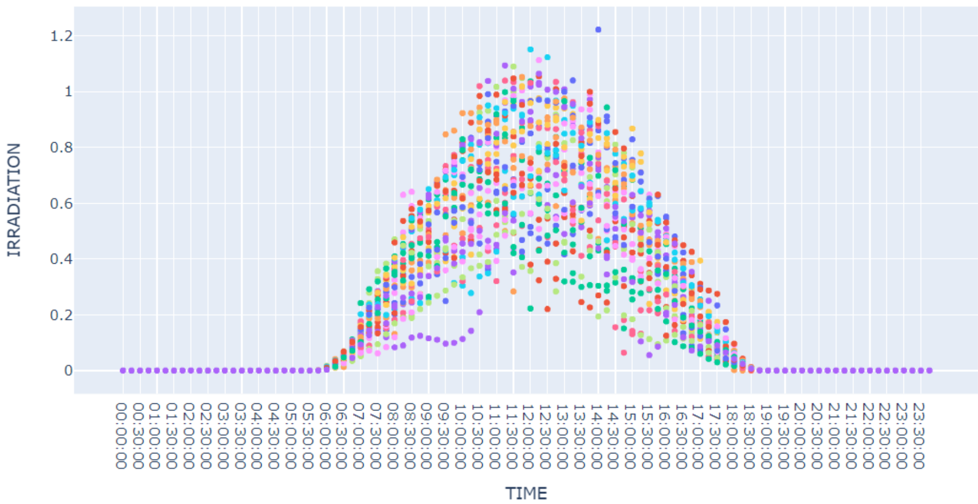

Fig. 33: Average Irradiation daily distribution

The Figure 32 reveals numerous instances where no power was generated during daylight hours, suggesting substantial disruptions in the energy production process. However, upon closer examination of the Figure 33, it becomes apparent that the irradiation levels never plummeted to sufficiently low levels during these periods of power loss. This discrepancy between the available sunlight and the lack of power generation strongly indicates equipment failure rather than environmental factors as the primary cause.

The consistent irradiation levels throughout the day, coupled with the intermittent absence of power generation, underscore the likelihood of malfunctioning equipment within our system.

Such anomalies cannot be attributed to variations in sunlight intensity or environmental conditions alone, as the irradiation levels remain consistently adequate for power generation.

This evidence points towards a critical need for thorough inspection and maintenance of our equipment to rectify the underlying issues causing these disruptions. Addressing equipment failures promptly is imperative to restore the efficiency and reliability of our solar energy infrastructure, ensuring uninterrupted power generation and maximizing our renewable energy output.

The ambient temperatures in our environment (Figure 34) fluctuate within a range of 20 to 35°C, reflecting typical variations in daily temperature. However, the temperatures of our photovoltaic modules exhibit a more significant range, spanning from 18 to 65°C. Notably, during daylight hours, the module temperatures tend to reach significantly higher levels compared to the surrounding air temperature. This phenomenon can be attributed to the absorption of solar radiation by the modules, which heats them up as they convert sunlight into electricity.

One interesting observation is the lag in ambient temperature compared to the daily cooldown of the modules. Despite fluctuations in ambient temperature, the modules tend to cool down at a faster rate than their surroundings. This lag can be attributed to several factors, including the thermal mass of the modules, the efficiency of heat dissipation mechanisms, and the insulation properties of the module materials. As a result, even after the ambient temperature begins to decrease, the modules may retain heat for a certain period before cooling down to match the ambient temperature.

Understanding these temperature dynamics is crucial for optimizing the performance and longevity of our photovoltaic system. By monitoring and managing module temperatures effectively, we can mitigate the impact of thermal stress on the modules, enhance energy production efficiency, and ensure the reliability of our solar energy infrastructure. Additionally, insights into temperature variations can inform maintenance schedules and help identify potential issues such as overheating or inadequate cooling mechanisms, allowing for timely interventions to maintain system performance.

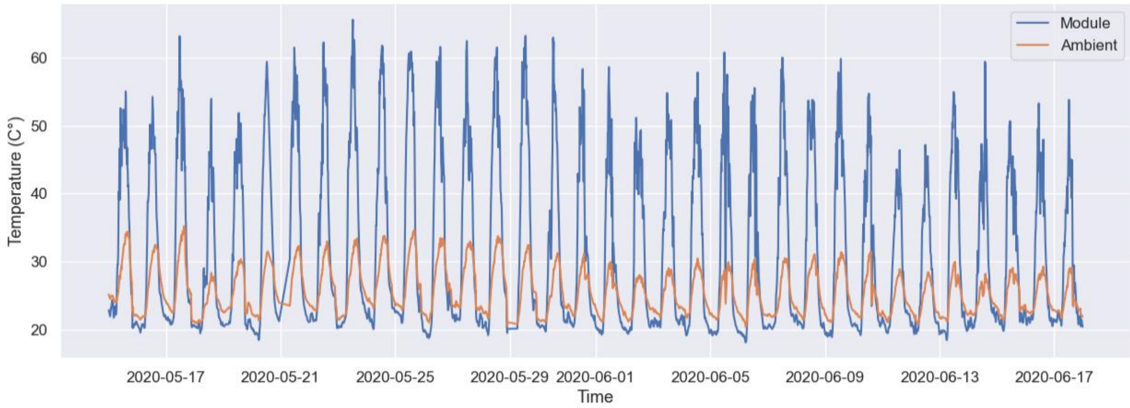

Fig. 34: Long-term Module and Ambient Temperature

Indeed, the presence of two days with significantly lower temperatures, indicative of "bad weather" conditions, underscores the challenges associated with weather forecasting, particularly in the absence of comprehensive weather data and advanced forecasting models. While ambient temperature fluctuations are one aspect of weather prediction, a multitude of other meteorological variables such as air pressure, wind speed and direction, humidity, cloud formation, and atmospheric stability play crucial roles in determining weather patterns.

Access to a broader range of weather data and the utilization of sophisticated forecasting models are essential for accurately predicting and anticipating adverse weather events. Incorporating variables such as air pressure, wind patterns, humidity levels, and cloud formations into forecasting models enables meteorologists to generate more reliable predictions of weather conditions, including temperature fluctuations and the likelihood of "bad weather" occurrences.

Furthermore, advanced weather forecasting techniques, such as numerical weather prediction models, utilize complex mathematical algorithms to simulate atmospheric processes and predict future weather conditions with greater accuracy. These models integrate vast amounts of observational data from satellites, weather stations, and other sources to generate detailed forecasts, including temperature variations and the probability of extreme weather events.

Enhanced access to comprehensive weather data and the implementation of advanced forecasting models are crucial steps toward improving the accuracy and reliability of weather forecasts, enabling individuals and organizations to better prepare for and mitigate the impacts of adverse weather conditions, such as temperature fluctuations associated with "bad weather" events.

4.5 Algorithm Implementation

A flowchart is a fundamental visual tool used to illustrate the operation of an algorithm in a clear and intuitive manner. It serves as a visual guide for understanding the flow of operations and decisions within a process or program. In the design and implementation of complex algorithms, a flowchart provides a schematic representation of operational phases, enabling developers to efficiently visualize and analyze the logical flow of instructions. This approach facilitates understanding, evaluation, and debugging of the algorithm, helping to reduce errors and improve overall efficiency of the system. Additionally, a flowchart provides a clear reference point for communication and collaboration among team members, enabling a shared understanding of the process and facilitating coordination of activities. In Figure 35 is illustrated the flowchart for the used algorithm.

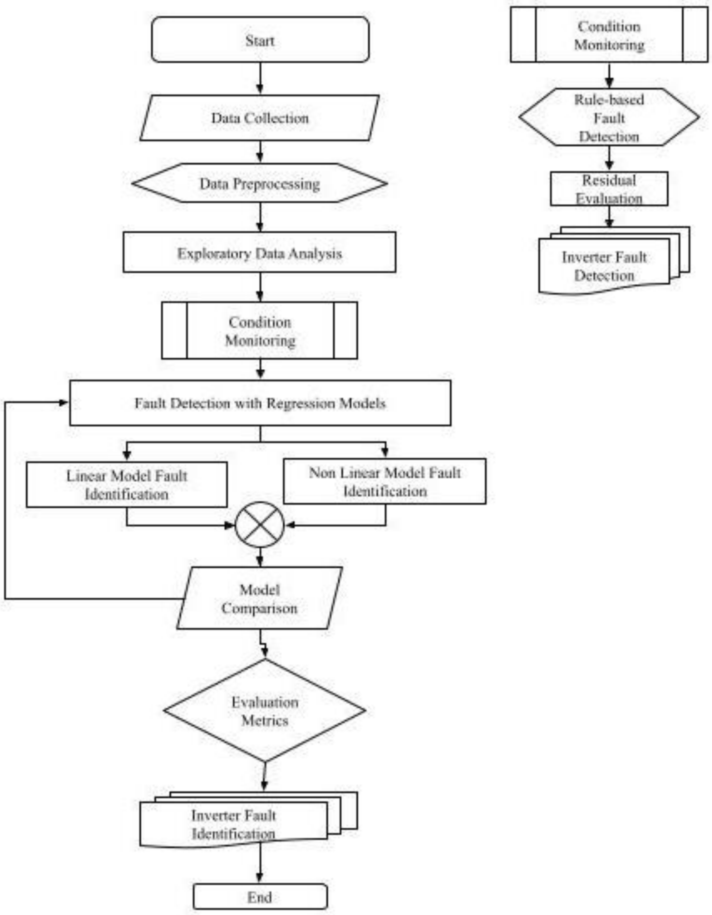

Fig.35: Algorithm flowchart

4.5.1 Condition Monitoring

Let's investigate the subroutine Condition Monitoring. Following the detection of evidence pointing towards possible equipment failure within our system, our focus now shifts towards implementing measures for automated equipment failure detection.

It is used a Rule-based Fault Detection:

Through our thorough data exploration process, we have identified a simple yet effective approach for detecting faulty equipment: namely, if there is no power recorded at the inverter during regular daytime operation, we can confidently infer equipment malfunction. To operationalize this detection method, we will proceed to create a new column labeled "STATUS" This column will serve as an indicator of faulty operation, allowing us to systematically flag instances of equipment failure within our dataset. By incorporating this rule-based fault detection mechanism, we aim to streamline the identification process and promptly address any equipment malfunctions, thus bolstering the reliability and performance of our system.

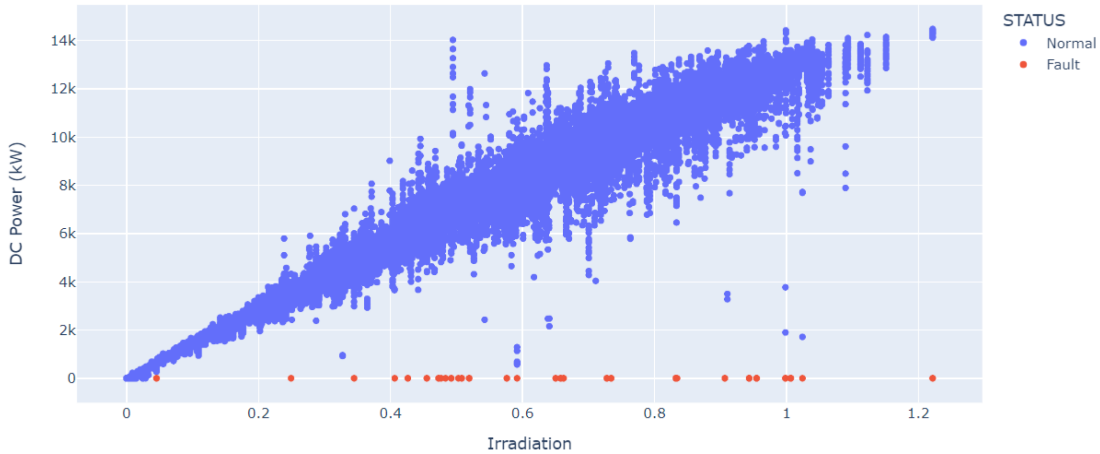

Fig.36: Fault detection in DC Power – Irradiation correlation

In the graph provided in Figure 37, the objective is to pinpoint the specific days and associated inverters where anomalies have been detected within our system. The classification system utilized here categorizes observations into two distinct groups: Normal and Fault. This binary classification simplifies the interpretation of the data, allowing for straightforward identification of instances where equipment operation deviates from the expected norm. By plotting the data on a graph with the y-axis representing the two possible configurations (Normal and Fault), we can visualize the distribution of anomalies across different days and corresponding inverters.

This visualization aids in identifying patterns or trends associated with equipment malfunction, facilitating targeted analysis and intervention to address underlying issues promptly. Furthermore, by clearly delineating between Normal and Fault states, the graph provides a concise overview of the prevalence and distribution of anomalies within our system, enabling effective decision-making and optimization of maintenance strategies.

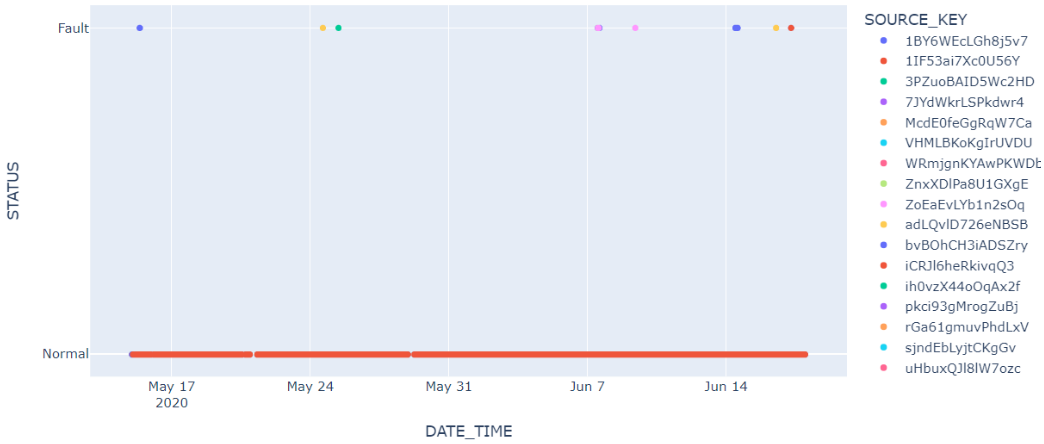

Fig. 37: Status of fault related to the inverters along the time

This type of analysis enables us to discern a total of 65 fault events associated with the unique identification number assigned to each individual inverter, as illustrated in Figure 38. By correlating the occurrence of faults with specific inverter identifiers, we gain valuable insights into the performance and reliability of each unit within our system. This detailed breakdown allows us to pinpoint the precise inverter units experiencing issues, facilitating targeted troubleshooting and maintenance efforts. Additionally, by visualizing the distribution of fault events across different inverter IDs, we can identify any patterns or clusters that may indicate systemic issues or recurring problems affecting multiple units. This comprehensive analysis serves as a crucial tool in optimizing the operational efficiency and longevity of our equipment, ensuring the continued reliability and performance of our system.

SOURCE_KEY	
bvBOhCH3iADSZry	21
1BY6WEcLGh8j5v7	19
z9Y9gH1T5YWrNuG	7
wCURE6d3bPkepu2	5
zBIq5rxdHJRwDNY	3
McdE0feGgRqW7Ca	3
sjndEbLyjtCKgGv	3
zVJPv84UY57bAof	2
ih0vzX44oOqAx2f	1
iCRJl6heRkivqQ3	1

Fig.38: Inverter ID and number of fault events

To conduct a Pareto Analysis evaluation (Figure 39), the fault events have been plotted on a bar chart sorted in descending order. This technique takes its name from the Italian economist Vilfredo Pareto and is based on the principle of the "vital few," where a small percentage of causes lead to the majority of observed effects. By arranging the fault events in this manner, it's possible to clearly visualize the frequency and distribution of fault occurrences across different categories or identifiers.

Pareto Analysis is a decision-making method used to identify and focus on the most significant factors contributing to a problem or failure. This approach is based on the idea that addressing the few critical factors can lead to significant improvements in the overall system. By identifying and addressing the primary contributors to the problem first, it's possible to maximize the effectiveness of intervention and optimize resource allocation.

The use of Pareto Analysis is widespread across various sectors and contexts, including industrial, commercial, and healthcare settings. It allows for the rapid identification of key inefficiencies or malfunctions, providing a starting point for implementing targeted corrective actions. Additionally, it helps focus attention on the areas that will yield the greatest impact, ensuring that resources are utilized efficiently to address the most pressing issues.

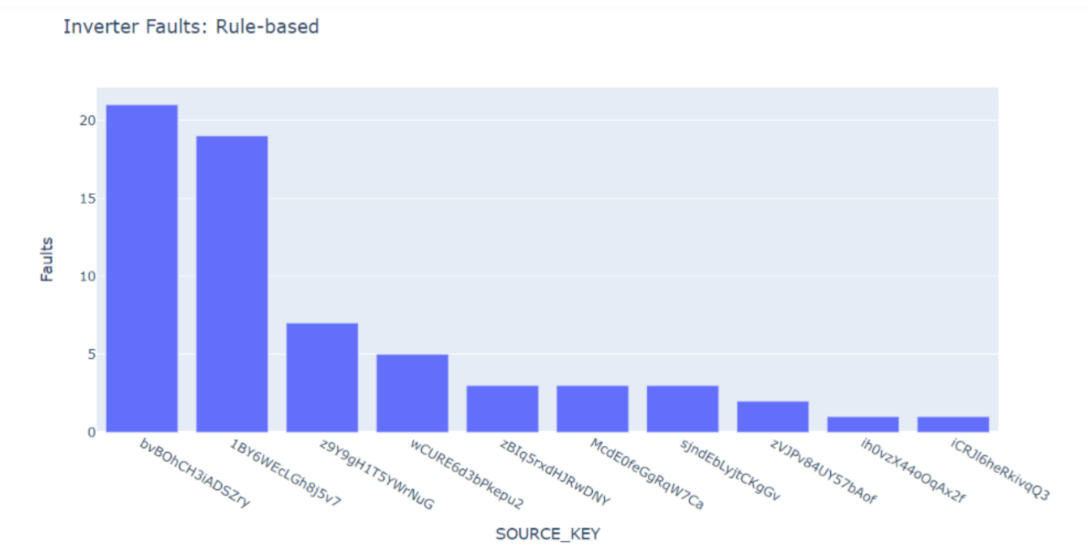

Fig. 39: Bar chart for Pareto analysis related to the inverters' fault

The analysis revealed that the highest number of faults occurred on two specific dates: June 7th, 2020, and June 14th, 2020. These dates stand out as periods with significantly increased occurrences of equipment malfunctions within the system. Moreover, upon examining the performance of individual inverters, it was observed that two particular units, namely bvBOhCH3iADSZry and 1BY6WEcLGh8j5v7, recorded the most failures compared to others. This insight is crucial for pinpointing specific components or units within the system that may be experiencing recurrent issues, thus necessitating focused attention and potential maintenance or replacement actions. Identifying these patterns of failure allows for targeted intervention strategies aimed at improving the reliability and overall performance of the system.

4.5.2 Resilience Engineering Framework Integration

In Figure 40, a detailed illustration of the integration with the Resilience Engineering framework is provided. This framework, originating from the field of complex systems engineering, focuses on analyzing and enhancing the resilience of operational systems, particularly in high-complexity and dynamic contexts such as systems and industrial process engineering.

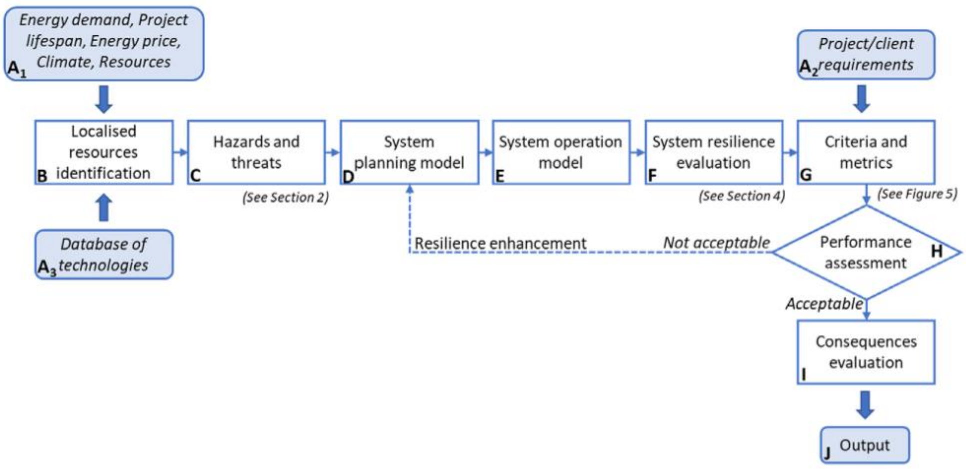

Fig.40: Resilience planning process, steps are labelled from A to J [5]

The integration with this framework represents a significant step in improving the robustness and adaptability of the system, as it allows for proactive identification and management of challenges and vulnerabilities that may arise during normal system operation.

Through the detailed analysis provided in the figure, it is possible to gain a better understanding of how the key principles and concepts of Resilience Engineering are applied and integrated within the specific context of the system under examination.

This integration enables the development of targeted strategies and procedures to effectively address crisis situations and ensure operational continuity even in the presence of unforeseen events or external stressors.

Furthermore, the utilization of the Resilience Engineering framework enables the adoption of a systemic and holistic approach to system safety and performance management, considering not only technological and infrastructural aspects but also organizational, human, and behavioral ones. This comprehensive approach allows for a more robust and resilient system design and operation, capable of effectively responding to and recovering from disturbances or disruptions.

While community-level energy systems, such as microgrids or district heating and cooling systems, appear to enhance energy resilience, there is a need for more explicit and consistent metrics in their planning, design, operation, and evaluation, particularly during disruption events. The transition of energy systems from centralized to decentralized, local, and community-level systems is underway. However, various hazards, both chronic and acute, can have short and long-term impacts on these systems, necessitating thorough analysis during energy master planning processes.

Understanding the dependencies, interdependencies, and coterminous effects of these systems is crucial as they can expose communities to high risks of performance disruption. Therefore, such scenarios should be thoroughly investigated in the early stages of community-level systems planning and design. To illustrate how the proposed resilience framework can be utilized, Figure 40 provides a flowchart outlining the main steps of resilience planning during energy master planning.

This flowchart demonstrates how resilience layers and metrics collaborate to evaluate various dimensions of resilience performance in a community-level energy system. Moreover, it showcases how resilience metrics can be incorporated into the planning, design, and operation of community systems to enhance their resilience. In practice, the energy master planning process may involve several additional steps and iterations, with varying levels of detail required at each design stage (conceptual, preliminary, and detailed design). The proposed resilience framework represents a significant advancement and complements existing resilience planning approaches, especially at the community level.

The initial steps of the resilience planning process, denoted as A1, A2, and A3 in Figure 40, involve gathering input data, client requirements, and technology/component information. Input data includes separate cooling, heating, and electricity demands for each node, energy prices, climate data, and other relevant information. Client requirements encompass project goals, limitations, and critical building energy requirements. The technology/components database provides details such as cost, lifespan, efficiency, flexibility, storage capabilities, temperature dependency, and more.

Subsequently, in step B, local resources and their availabilities, such as energy networks, wind, solar, and geothermal potentials, are identified. Step C involves determining threats to the local community energy system and the grid infrastructure, including hazards frequency, probability, magnitude, and impacts.

In step D, the system is planned and modeled, encompassing system architecture, component sizing, configuration, and placement. Resilience design guidelines, frameworks, and recommended attributes are utilized in this stage, along with considerations for scenarios representing mission-critical uses under disruption conditions.

The modeling approach is chosen based on the project design phase, considering both new developments and existing upgrades. The proposed resilience framework and dimensions guide the planning and design process.

Step E involves modeling system operation and operation strategies throughout the community system's lifespan. In step F, system resilience performance is evaluated across three resilience layers: engineering-designed resilience, operational resilience, and community resilience. This evaluation considers various factors such as system design, operational strategies, adaptive behaviors, and energy use reduction potential.

Each resilience layer contributes differently to overall system resilience. Engineering-designed resilience provides the first line of defense, while community resilience has the most significant impact on overall resilience capacity. Operational resilience is activated after reaching the maximum engineering-designed resilience. The duration of each resilience phase depends on both internal system features and external factors, such as hazard characteristics and available resources.

Consideration of aging effects and degradation profiles is essential in understanding the dynamic nature of energy performance recovery. The resilience trapezoid model depicts four main temporal phases: robust state, after the event state, recovery state, and post-recovery state. The duration of each phase varies based on system features and external factors impacting restoration time. The resilience of energy infrastructure and its response to disruptions can vary based on the type of infrastructure and the nature of the disruption. In some cases, such as acute operational threats, the operational resilience capacity may be compromised before or simultaneously with the engineering-designed resilience capacity. However, in most scenarios, there are intersections within resilience layers and metrics, with interrelated phases within resilience layers.

Resilience layers may intersect, and metrics may potentially estimate the combined capacity of engineering-designed and operational resiliencies.

Assessing community-level energy resilience involves employing or extending existing frameworks, methods, and models. These assessment methods can be categorized based on the level of data required, modeling approach, and methodology.

Criteria, indicators, and metrics are selected in Step G based on decision-maker goals. Theoretical performance-based metrics for assessing energy resilience at the community level are illustrated in Figure 41. These metrics are categorized into four phases: prepare, withstand, adapt, and recover.

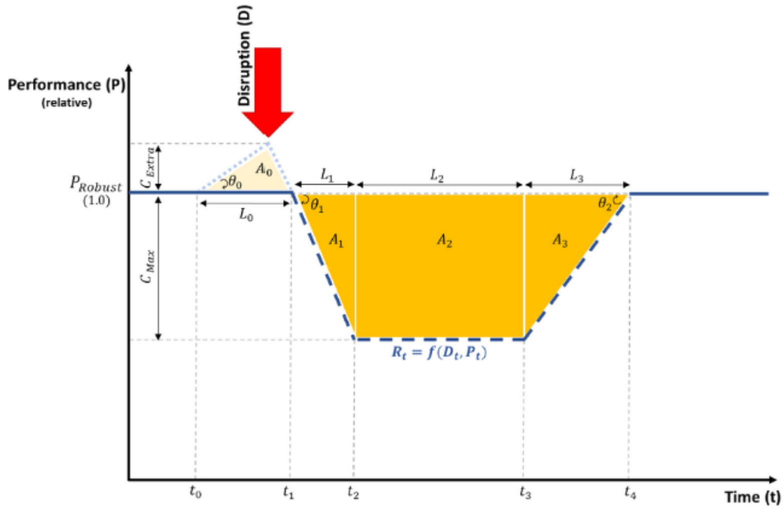

Fig.41: Resilience metrics for community-level energy master planning [5]

Metrics	Unit	Phases			
		Prepare	Withstand	Adapt	Recover
Resilience length	hours	$L_0 = t_1 - t_0$	$L_1 = t_2 - t_1$	$L_2 = t_3 - t_2$	$L_3 = t_4 - t_3$
Rate of resilience	performance/hours	C_{Extra}/L_0	$\tan(\theta_1) = C_{Max}/L_1$	–	$\tan(\theta_2) = C_{Max}/L_3$
Resilience area	performance-hours	A_0	A_1	A_2	A_3
Resilience capacity	performance	C_{Extra}	C_{Max}	C_{Max}	–
Total resilience length	hours	$L = t_4 - t_0 = L_0 + L_1 + L_2 + L_3$			
Total resilience area	performance-hours	$A = A_0 + A_1 + A_2 + A_3$			
Total resilience capacity	performance	$C = C_{Extra} + C_{Max}$			

Table 9: Summary of theoretical resilience metrics in each resilience phase

Preparation mechanisms are typically employed before an extreme event, with the time and extra resilience capacity depending on prediction ability and event type. During the withstand phase, the system performance initially withstands the threat until reaching its maximum resilience capacity. The system then adapts to the new conditions until the recovery phase begins.

Theoretical resilience metrics can be utilized across resilience layers to estimate system resilience and related consequences, such as economic loss. Design choices depend on decision-makers' objectives and the impacts of performance disruptions.

Constraints such as limited economic or environmental resources can affect achieving the desired level of energy resilience.

System resilience can be enhanced by minimizing the area of the resilience trapezoid, and the performance of each design alternative should be assessed against the required level of service performance. If acceptable, the process continues to subsequent steps; otherwise, additional design alternatives with enhanced resilience should be considered and the process repeated.

Depending on decision-makers' objectives, community-level energy systems can be designed differently, considering trade-offs in resilience capacity, resilience slopes, and recovery time. However, current practices may face limitations due to a lack of evidence-based knowledge for designing and implementing effective intervention options.

During energy master planning for communities, various measures can be implemented to enhance energy resilience across different layers and phases. For instance, Demand Side Management (DSM) can contribute to both operational and community resilience. When facility management adjusts operational logic, such as changing set-point temperatures, without involving residents, it serves as an operational measure. Conversely, when residents actively adjust their energy consumption post-disruption to adapt to the system's new state, it becomes a community-societal resilience measure. Some measures, like energy storage and smart operations, play essential roles across multiple resilience phases.

These resilience enhancement measures can be categorized into short-term, such as prediction, monitoring, and priority setting, and long-term measures, including physical upgrades and energy storage. When assessing energy resilience, it's essential to consider not only system performance but also the cost and emissions associated with alternative solutions. Systems with lower resilience costs, which encompass systemic impact and total recovery effort, are deemed more resilient to disruptions.

Resilience metrics are pivotal in decision-making during energy master planning to evaluate engineering performance and consequences related to disruption, such as economic, environmental, and social impacts. These consequences are directly linked to the resilience trapezoid, including its area, slopes, depths, and lengths. Resilience and sustainability should be viewed as complementary attributes of the system during design and operation. While sustainability impacts are typically considered during standard design and operation, resilience consequences arise during unexpected extreme event conditions.

However, trade-offs may arise among design options concerning their sustainability and resilience performances.

The proposed resilience framework facilitates a performance-based assessment of alternative community system designs and their sustainability-related consequences, allowing for a more comprehensive evaluation of community system performance.

The output of this process enables decision-makers to compare and select design scenarios based on resilience performance (and consequences), potentially incorporating other performance and sustainability metrics. Each resilience layer and relevant metrics can be weighted separately according to decision-makers' preferences. For instance, when facing limited capital funding, decision-makers may prioritize economic consequences and assess trade-offs among resilience layers or phases.

The systematic application of energy resilience at the community level is not yet common practice, but insights can be gleaned from similar approaches and energy systems.

4.5.3 Fault Detection with Regression Models

As previously discussed in 4.1.1, the fault detection with regression models is a pivotal aspect of predictive maintenance and quality control in various industries. Leveraging the power of regression analysis, these models aim to identify anomalies or deviations from expected patterns within a system or process. By analyzing historical data and establishing relationships between input variables and potential faults, regression models offer a proactive approach to detecting abnormalities before they escalate into critical issues. This paragraph sets the stage for exploring the intricate intersection of regression analysis and fault detection, highlighting its significance in ensuring operational efficiency and product reliability.

4.5.3.1 Linear Regression Model

In this initial exploration of predicting DC power from Irradiance, I embarked on a simplistic linear model, encapsulated by the equation:

$$P(t) = a + bE(t) \quad (20)$$

Here, $P(t)$ denotes the generated DC power at time t, while $E(t)$ represents the irradiance observed at the same time instant. The parameters a and b serve as coefficients of the linear model, aiming to capture the presumed linear correlation between irradiance and DC power output.

However, the operational dynamics of PV modules extend beyond linear approximations, necessitating a deeper investigation into the interplay between irradiance, temperature and DC power generation. A critical consideration lies in the temperature dependence of PC cell efficiency, which can significantly impact overall performance. Elevated temperature, often exceeding 65°C in operational settings, include various effects, including increased carrier recombination rates and decreased open-circuit voltage, thereby diminishing the overall efficiency of PV cells.

To accurately model this temperature dependency and its influence on DC power generation, we turn to the diode equation, a fundamental equation in semiconductor physics governing the behavior of PN junctions:

$$I_{PV} = I_L - I_0\left(e^{\frac{q(V_{OC}+m\Delta T)}{kT}} - 1\right) \quad (21)$$

In the equation (21), I_{PV} represents the PV cell current, I_L denotes the light-generated current, I_0 signifies the diode reverse saturation current, V_{OC} symbolizes the open-circuit voltage, m reflects the temperature coefficient, ΔT signifies the temperature deviation from the reference temperature, and q, k and T represent the elementary charge, Boltzmann constant, and absolute temperature respectively.

By incorporating the temperature-dependent behavior of PV cells into the model and employing advanced nonlinear regression techniques in the next paragraph, I aim to refine the predictive capabilities and provide deeper insights into the complex relationship between irradiance, temperature, and DC power output in PV systems. This comprehensive approach promises to enhance the accuracy and reliability of fault detection algorithms in real-world applications, thereby contributing to improved efficiency and performance in solar energy systems.

Fig. 42: Fault identification result from linear model

To evaluate the occurrence of fault detection instances (Figure 42), I engaged in an in-depth analysis of residuals, defined as the discrepancy between the measured DC power and the DC power estimated by the regression model. This analysis involved scrutinizing the temporal evolution of residuals to detect any deviations indicative of anomalies or faults within the system. Formally the residual $\rho(t)$ at time t is expressed as the difference between the measured DC power $P_m(t)$ and the predicted DC power $P_p(t)$ derived from the regression model:

$$\rho(t) = P_m(t) - P_p(t) \quad (22)$$

This quantifies the discrepancy between observed and model-predicted values at each time point. By examining the statistical properties, such as mean, variance, and autocorrelation, of the residual time series, we can discern patterns indicative of underlying faults or abnormalities. The identification of significant deviations in the residual signal provides valuable insights for proactive maintenance and fault diagnosis, thereby enhancing the operational reliability and efficiency of the system.

Several academic references underscore the significance of residual analysis in fault detection methodologies. For instance, Khan et al. (2021) [32] provide a comprehensive overview of fault detection techniques in photovoltaic systems, emphasizing the role of residual analysis in anomaly detection. Similarly, Wang et al. (2020) [33] discuss the importance of residual analysis in data-driven fault detection and diagnosis methodologies. These studies highlight the critical role of residual analysis in identifying and diagnosing faults, thereby enhancing the reliability and efficiency of the system. Additionally, Fathi et al. (2021) [34] offer further evidence of the value of residual analysis in the context of photovoltaic modules, underscoring its utility as an essential component of anomaly diagnosis strategies.

In Figure 34, the graph illustrates the behavior of DC power, showcasing data points obtained from actual measurements, predictions made by the model, and the corresponding residuals.

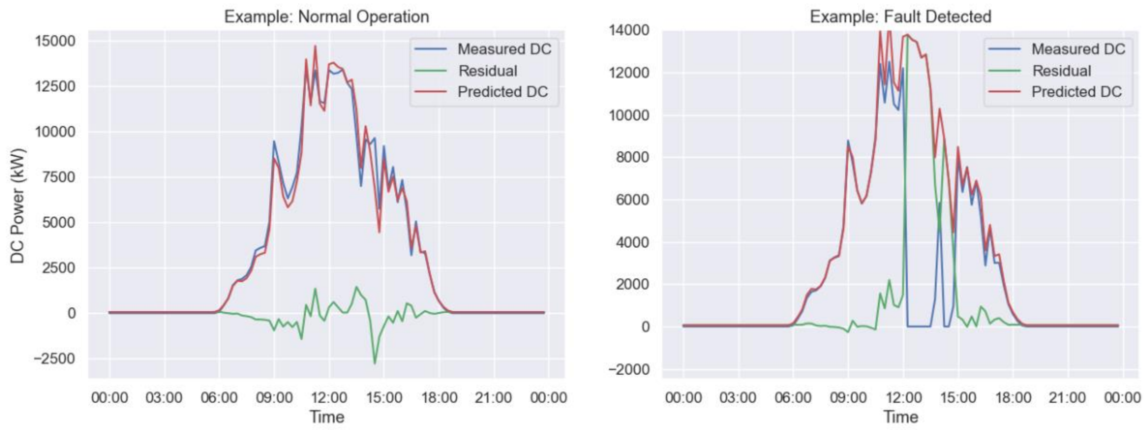

Fig.43: DC Power Normal and Fault operations evaluated through the linear model residual computation

When the residual value is zero, it indicates a perfect match between the measured and predicted DC power values, signifying no discrepancy between the two. Conversely, if the residual value is non-zero but falls within an acceptable range, it suggests that the difference between the measured and predicted values is within an acceptable margin. However, when the residual peaks, indicating a scenario where the measured value is zero and aligns with the predicted value, it signals the presence of an anomaly or abnormality in the system. This observation is critical for fault detection purposes, as it provides insights into the system's performance and highlights potential areas of concern without the need for explicit mathematical formulations or references.

Fig. 44: Fault detection evaluation along the entire dataset for the inverter 22

In Figure 44, the model is applied to the entire dataset to highlight the detection of a fault over the entire sampled time span.

This comprehensive application of the model across the dataset enables the visualization and identification of any fault occurrences or abnormalities that may have occurred throughout the sampled temporal range. This approach allows for a holistic assessment of the system's performance and fault detection capabilities over time without the need for segmented analysis.

4.5.3.2 Non-Linear Regression

According to Hooda et al. (2018) [31], the generated power $P(t)$ of a PV cell at time t can be modeled by the nonlinear equation:

$$P(t) = aE(t)\left(1 - b\left(T(t) + \frac{E(t)}{800}(c - 20) - 25\right) - dln\big(E(t)\big)\right) \quad (23)$$

Where:

- $P(t)$ represents the generated power of the photovoltaic cell at time t. It is influenced by both irradiance and temperature.
- $E(t)$ denotes the irradiance at time t, which is a crucial factor affecting the power output of the photovoltaic cell. Irradiance represents the intensity of solar radiation incident on the cell’s surface.
- $T(t)$ represents the temperature of the PV cell al time t. Temperature fluctuations can significantly impact the performance of PV cells, affecting their efficiency and power output.
- Coefficients a, b, c and d:
  - a determines the overall scaling factor for the power output
  - b modulates the impact of temperature and irradiance on the power generation process. It adjusts the response of the cell to changes in temperature and irradiance.
  - c represents a coefficient that scales the temperature term within the equation, affecting how temperature influences the power output.
  - d modifies the logarithmic term, capturing the nonlinear relationship between irradiance and power generation.

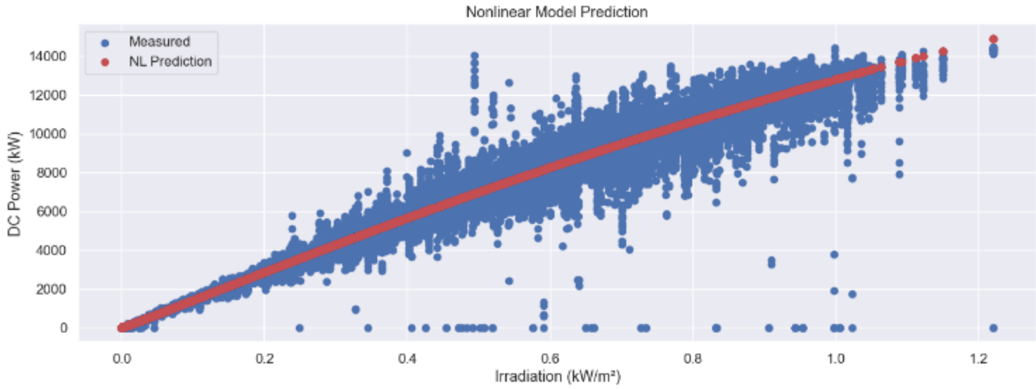

Fig. 45: Nonlinear model prediction of DC Power

The graph in Figure 45 illustrates the prediction of the nonlinear model for DC power concerning solar irradiation. This graph provides a visualization of the anticipated relationship between the incident solar irradiance on the photovoltaic cell and the power generated by it in DC.

The curve depicted in the graph follows a nonlinear trend, as anticipated by the model proposed by Hooda et al. (2018). [31] Generally, an increase in solar irradiance corresponds to an increase in generated power, aligning with the operational characteristics of photovoltaic systems. However, the curve's trajectory is nonlinear due to influences from factors such as temperature and the nonlinear effects of irradiance on power generation, as considered in the model.

This graph offers a visual prediction of the expected power output from the photovoltaic cell in response to variations in solar irradiance. It serves as a valuable tool for understanding how the nonlinear model accounts for various variables and interactions to accurately predict the power generated by the photovoltaic cell under different environmental conditions.

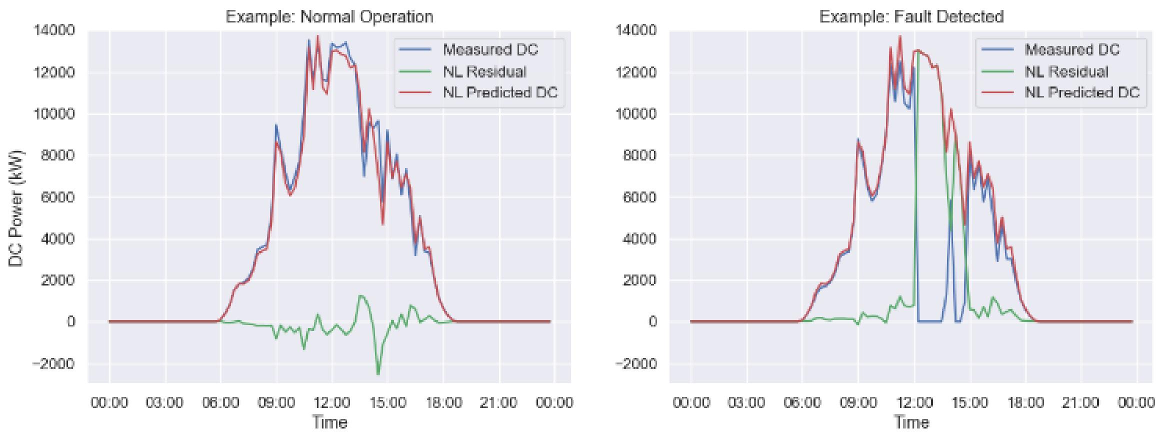

Fig.46: DC Power Normal and Fault operations evaluated through the nonlinear model residual computation

The graphs depict DC Power under normal and fault operations, evaluated through the computation of residuals using the nonlinear model in equation (23).

These graphs provide a visual representation of the difference between the predicted DC power output by the nonlinear model and the actual measured DC power. Positive residuals indicate overestimation by the model, while negative residuals indicate underestimation. By comparing the residual patterns under normal and fault operations, insights can be gained into the performance of the photovoltaic system. Anomalies or faults may be indicated by significant deviations between the measured and predicted DC power, as reflected in the residuals.

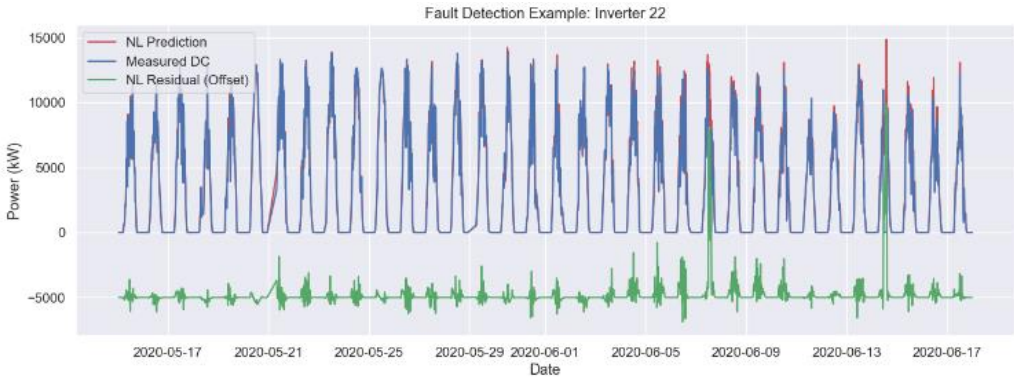

Fig. 47: Fault detection evaluation along the entire dataset for the inverter 22 with nonlinear model

The figure 47 provides a comprehensive assessment of fault detection across the entire dataset specifically focusing on inverter 22, leveraging the capabilities of the nonlinear model. This evaluation serves as a crucial step in monitoring and maintaining the performance of photovoltaic systems, particularly in identifying any irregularities or malfunctions that may occur over an extended period. By applying the nonlinear model to the entire dataset, the analysis captures a holistic view of the system's behavior and potential fault occurrences. This approach enables proactive identification of anomalies, allowing for timely intervention and corrective actions to be taken to ensure optimal system efficiency and reliability. The visualization presented in Figure 47 facilitates a detailed examination of fault detection outcomes, providing valuable insights into the effectiveness of the nonlinear model in detecting and diagnosing faults in inverter 22 operations.

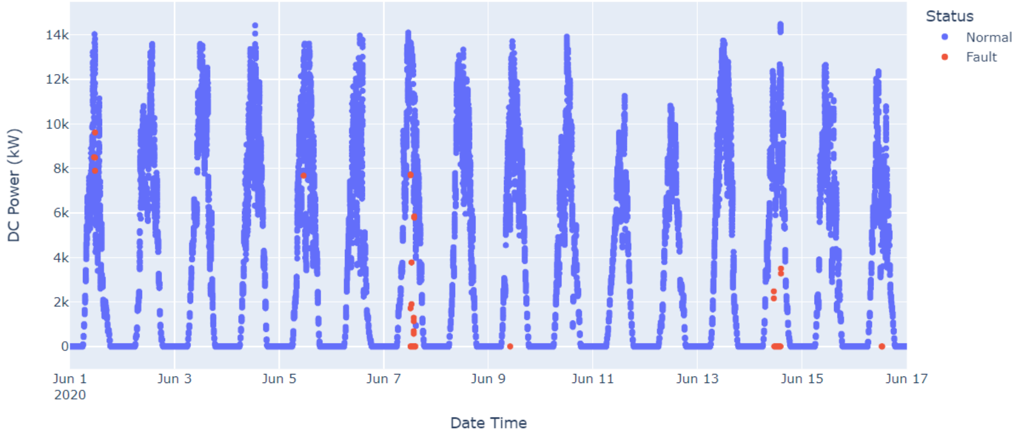

Fig. 48: Underperformance and faults analysis using nonlinear model

The graph in Figure 48 presents a visual analysis of underperformance and fault occurrences in the DC power output of a photovoltaic system over time, as predicted by a nonlinear model.

The x-axis represents the datetime, indicating the time period during which the data was collected or analyzed. Meanwhile, the y-axis represents the DC power output of the photovoltaic system, measured in $[kW]$.

The plot likely includes multiple lines or markers representing different aspects of the analysis. These may include:

1. DC Power: The DC power output of the photovoltaic system based on the nonlinear model. This line or curve illustrates the expected performance of the system under normal operating conditions.

2. Underperformance Events: These could be indicated by points or markers where the DC power falls. Underperformance events may indicate issues such as suboptimal system operation or environmental factors affecting performance.

3. Fault Events: These might be represented by distinct markers or annotations highlighting instances where the measured DC power deviates significantly from the predicted values, indicating potential faults or anomalies in the system.

Overall, the graph provides a comprehensive overview of the photovoltaic system's performance over time, enabling analysts to identify periods of underperformance and potential fault occurrences. This information is valuable for diagnosing issues, optimizing system operation, and improving overall system resilience and reliability.

4.6 Results and Interpretation

Comparing a linear model to a nonlinear one for fault detection in off-grid PV systems serves several crucial purposes.

1. Prediction Accuracy: Nonlinear models can better capture the complex relationships among variables compared to linear models. In fault detection, where small data variations may indicate anomalies or faults, a nonlinear model might be more accurate in detecting such variations, enabling a more precise diagnosis of issues in the PV system.

2.Understanding Phenomena: Comparing the two models can aid in better understanding the phenomena influencing PV system behavior. For instance, if the nonlinear model shows significant discrepancies from the linear model, it may indicate the presence of nonlinear effects in the system, such as responses that are not proportional to input variable changes. This deeper understanding can guide fault detection efforts and issue resolution.

3. Robustness and Adaptability: While nonlinear models may offer greater accuracy, they can also be more complex and resource-intensive compared to linear ones. However, fault detection in an off-grid PV environment can benefit from more sophisticated models capable of adapting to environmental and load variations. Assessing the trade-offs between accuracy, complexity, and computational resources is essential for developing effective fault detection strategies tailored to off-grid PV systems.

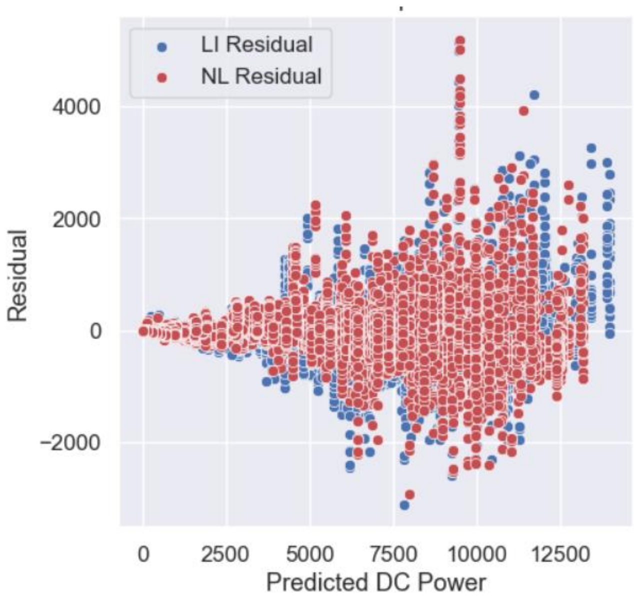

Fig. 49: Model comparison between linear and nonlinear model

Comparing a linear model to a nonlinear one (Figure 48) for fault detection in off-grid PV systems is particularly valuable from the perspective of Resilience Engineering. Resilience Engineering emphasizes the system's ability to adapt and recover from disruptions, including faults and anomalies.

The comparison is useful in particular to enhance the Fault Detection, to understand the system dynamic and to develop adaptive strategies.

Enhanced Fault Detection because a nonlinear model often exhibits greater sensitivity to subtle changes and nonlinear relationships within the system. By comparing the outputs of linear and nonlinear models, we can identify discrepancies that may indicate potential faults or anomalies not detected by the linear model alone. This enhanced fault detection capability contributes to the system's resilience by enabling the timely identification and mitigation of emerging issues.

Understanding System Dynamics because comparing linear and nonlinear models provides insights into the underlying dynamics of the PV system. Nonlinearities in the system's behavior, such as nonlinear responses to environmental changes or load fluctuations, can be better captured and understood through the nonlinear model. This understanding facilitates proactive measures to enhance system resilience, such as adapting operational strategies to account for nonlinear effects.

Adaptive Strategies because Resilience Engineering advocates for adaptive strategies that can respond effectively to changing conditions and unexpected events. By leveraging the insights gained from comparing linear and nonlinear models, operators can develop adaptive fault detection and response strategies. These strategies can incorporate both linear and nonlinear aspects of system behavior, allowing for more robust and flexible responses to disturbances and uncertainties.

The implementation of the algorithm underpinning the fault detection study presented in this thesis is beneficial during both the community master planning phase and operational phases for assessing the proper functioning of the system and consequently the microgrid. Integrating it with the framework of Resilience Engineering allows for the development of a more resilient and self-sufficient energy system from the early stages, particularly in off-grid contexts.

During the community master planning phase, the algorithm aids in designing a robust energy infrastructure by identifying potential faults and vulnerabilities. By considering fault detection mechanisms early in the planning process, communities can implement proactive measures to mitigate risks and enhance the resilience of their energy systems.

In the operational phase, the algorithm continuously monitors the performance of the system, promptly detecting any deviations from expected behavior.

This real-time fault detection capability is crucial for ensuring the reliable operation of the microgrid and minimizing downtime. Additionally, by integrating with the resilience engineering framework, the system can dynamically adapt to changing conditions, effectively managing disruptions and optimizing energy self-sufficiency.

Overall, the implementation of this fault detection algorithm contributes to the development of resilient and self-sustaining energy systems, particularly in off-grid settings where reliability and autonomy are paramount. By leveraging the principles of resilience engineering, communities can build energy infrastructures that are not only capable of withstanding disruptions but also thrive in challenging environments.

4.7 Limitation and potential improvements

4.7.1 Weather limitation and improvements

The implementation of a fault detection algorithm in a microgrid, while beneficial, may encounter limitations, particularly when dealing with external factors such as weather conditions. For instance, consider the scenario where there are two days characterized by significantly lower temperatures, indicative of "bad weather" conditions. Such events can pose challenges for fault detection algorithms, as they may impact the performance of the microgrid in unpredictable ways.

The difficulty arises from the lack of comprehensive weather data and advanced forecasting models. While the fault detection algorithm may be able to monitor internal parameters of the microgrid, such as voltage levels or power output, it may not have access to detailed weather data, including air pressure, wind speed, humidity, cloud formation, and other relevant factors. These weather variables can significantly influence the behavior of renewable energy sources, such as solar panels and wind turbines, as well as the overall energy demand within the microgrid.

Without access to more extensive weather data and sophisticated forecasting models, the fault detection algorithm may struggle to accurately predict and account for the impact of adverse weather conditions on the microgrid's performance. As a result, faults or anomalies triggered by weather-related factors may not be promptly detected or appropriately addressed, potentially leading to disruptions in the microgrid's operation.

To mitigate this limitation, it may be necessary to integrate the fault detection algorithm with external weather monitoring systems or advanced weather forecasting models. By incorporating real-time weather data into the analysis, the algorithm can better anticipate and respond to weather-induced fluctuations in energy generation and demand within the microgrid.

Additionally, employing adaptive strategies informed by weather forecasts can help enhance the resilience of the microgrid, enabling it to withstand and recover from adverse weather events more effectively.

4.7.2 Linear Model limitation and improvements

Using a linear regression model in a fault detection algorithm for a microgrid may have several limitations. Firstly, linear regression assumes a linear relationship between the input variables and the output, which may not accurately capture the complex and nonlinear dynamics of a microgrid system, especially during fault conditions [37]. Microgrids often exhibit nonlinear behaviors due to the presence of renewable energy sources, energy storage systems, and dynamic loads, which can lead to inaccuracies in fault detection when using linear models [35]. Additionally, linear regression may struggle to detect faults that occur abruptly or exhibit transient behavior, as it relies on historical data to predict future outcomes [39]. Furthermore, linear regression assumes that the input variables are independent of each other, which may not hold true in the context of microgrid operations where various parameters are interrelated [40]. Finally, linear regression is sensitive to outliers in the data, which can distort the model's predictions and lead to false fault detection alarms [36]. To address these limitations, alternative fault detection methods such as machine learning algorithms (e.g., support vector machines, artificial neural networks) or hybrid models that combine linear regression with nonlinear techniques may be considered [38].

In order to mitigate the limitations associated with using a linear regression model for fault detection in microgrids, various improvements and alternative approaches can be considered. Firstly, incorporating more advanced machine learning algorithms, such as support vector machines (SVMs) or artificial neural networks (ANNs), can enhance fault detection accuracy by capturing nonlinear relationships and complex system dynamics more effectively [41]. These algorithms have the capability to learn from large datasets and adapt to changing operating conditions, making them suitable for fault detection in microgrids with diverse and evolving characteristics. Additionally, hybrid models that combine linear regression with nonlinear techniques, such as fuzzy logic or genetic algorithms, can leverage the strengths of each approach while mitigating their respective limitations [36].

Furthermore, incorporating real-time data streams from sensors and advanced monitoring systems into fault detection algorithms can improve their responsiveness and accuracy, enabling quicker detection and isolation of faults before they escalate into larger disruptions [42].

By integrating these advancements into fault detection algorithms, microgrid operators can enhance the reliability, resilience, and overall performance of their systems, ensuring continuous and efficient energy supply to end-users.

5. Conclusions

In conclusion, fault detection in microgrids is a critical aspect of ensuring the reliability and stability of these complex energy systems. While linear regression models offer simplicity and interpretability, they may not adequately capture the nonlinear and dynamic behaviors present in microgrid operations. As demonstrated, alternative approaches such as advanced machine learning algorithms and hybrid models can provide more accurate and robust fault detection capabilities by accommodating complex system dynamics and interdependencies. Furthermore, integrating real-time data streams and advanced monitoring systems enhances the responsiveness and effectiveness of fault detection algorithms, enabling prompt identification and isolation of faults to prevent system disruptions. Moving forward, continued research and development efforts in fault detection methodologies are essential to address the evolving challenges and complexities of microgrid operations, ultimately contributing to the advancement of sustainable and resilient energy infrastructure.

5.1 Summary of main results

The research conducted in this thesis has yielded several important findings in the field of resilience engineering for minigrids. Firstly, through a comprehensive literature review, key concepts and frameworks in energy access, minigrid design, resilience engineering, exploratory data analysis (EDA), and anomaly detection were synthesized and contextualized within the study's scope. This literature review provided a solid foundation for understanding the challenges and opportunities in enhancing the resilience of minigrid systems.

Building upon this foundation, the methodology section outlined the study context and described the design and implementation of the Resilience Engineering Framework (REF), which aimed to integrate anomaly detection techniques into minigrid resilience assessment and enhancement processes. The hybrid REF - Anomaly Detection Algorithm developed in this study combined resilience indicators with anomaly detection algorithms to identify and mitigate potential disruptions in minigrid operations.

The results obtained from the application of the integrated framework demonstrated its effectiveness in identifying anomalies and enhancing the resilience of minigrid systems. Through exploratory data analysis and algorithm implementation, the hybrid REF - Anomaly Detection Algorithm successfully detected and interpreted anomalies in minigrid performance data, providing valuable insights for system optimization and risk management.

The analysis and discussion section critically evaluated the implications of the study's findings, both theoretically and practically. It highlighted the significance of integrating anomaly detection into resilience engineering frameworks for minigrids and discussed potential applications and limitations of the hybrid approach.

This thesis has contributed to advancing knowledge and understanding in the field of minigrid resilience engineering by developing and implementing a novel framework that integrates anomaly detection techniques. While the study has achieved significant results, it also acknowledges its limitations and identifies potential avenues for future research and development in this area.

5.2 Potential Future Developments

Looking ahead, there are several potential avenues for future research and development in the field of minigrid resilience engineering and anomaly detection.

Firstly, further refinement and optimization of the hybrid REF - Anomaly Detection Algorithm could enhance its effectiveness in detecting and mitigating anomalies in minigrid operations. This could involve exploring alternative anomaly detection techniques, such as deep learning algorithms or ensemble methods, to improve the accuracy and robustness of anomaly detection models [27]. Additionally, expanding the scope of the framework to incorporate real-time monitoring and predictive analytics capabilities could enable proactive risk management and preemptive maintenance strategies, further enhancing the resilience of minigrid systems.

Moreover, future research efforts could focus on addressing the unique challenges and constraints faced by minigrids in different geographical and socio-economic contexts.

This could involve adapting the resilience engineering framework to suit the specific needs and characteristics of diverse minigrid deployments, such as remote rural areas, urban slums, or refugee camps. Furthermore, exploring innovative financing mechanisms and business models for minigrid resilience enhancement projects could help overcome barriers to implementation and scale-up, ensuring sustainable and equitable access to reliable energy services for underserved communities [43].

Additionally, advancing data analytics techniques and sensor technologies could facilitate more comprehensive and accurate anomaly detection in minigrids. This could include leveraging advanced data fusion techniques to integrate heterogeneous data sources, such as satellite imagery, weather forecasts, and IoT sensor data, for holistic minigrid monitoring and analysis. Furthermore, exploring the potential of blockchain technology and decentralized energy management systems could enhance the transparency, security, and efficiency of minigrid operations, enabling peer-to-peer energy trading and decentralized decision-making processes.

Moreover, fostering interdisciplinary collaboration and knowledge exchange among researchers, practitioners, policymakers, and local communities could facilitate co-design and co-implement innovative solutions for minigrid resilience enhancement. This could involve participatory approaches, such as co-creation workshops, community-based monitoring, and citizen science initiatives, to ensure that resilience interventions are contextually appropriate, socially inclusive, and environmentally sustainable (Brew-Hammond, 2010).

In conclusion, the future of minigrid resilience engineering and anomaly detection holds great promise for advancing energy access, sustainability, and resilience in off-grid and underserved communities. By harnessing emerging technologies, fostering interdisciplinary collaboration, and prioritizing community engagement and empowerment, we can unlock the full potential of minigrids as catalysts for inclusive and sustainable development.

6. Acknowkedgmentes

Ai miei genitori che con pazienza e tatto mi hanno supportato in un lungo cammino.

Alle mie sorelle che non hanno mai lesinato di essere vittime dei miei scherzi. Forse in tal senso loro dovrebbero ringraziare me.

A Francesca che con la sua empatia ha ben capito quando era momento di chiedere e quando di ascoltare.

Ai miei amici, parti del mio cuore e parti di me, ovunque io e loro siano. Lorenzo, Jacopo, Riccardo, Alessia.

Ai miei amici e colleghi in questo cammino ad elevata rugosità: Marcolino, Luca, Melania, Francesco, Ashiq, Benedetta. Meglio di chiunque altro avete capito ogni passaggio.

Ai miei nonni, il cui ricordo guida sempre il mio cammino.

Al Professor Micangeli, mentore e guida in un mondo che troppo spesso porta a perderci e a perdere ciò che è veramente utile.

Ai compagni del gruppo FSA con cui ho condiviso momenti indimenticabili in giro per il mondo e di fronte ad un bando da scrivere all'ultimo minuto. A Daniele, amico e maestro.

7. Bibliography

[1] International Energy Agency (IEA), (2010) *Energy Access: How to make modern energy access universal.*

[2] Knowledge Note of CIF Climate Investment Fund, (2014), *Increasing Rural Energy Access through Mini-Grids*

[3] Subhes C. Bhattacharyya (2012), *Energy access programmes and sustainable development: A critical review and analysis*, Energy for Sustainable Development Volume 16, Issue 3, September 2012, Pages 260-271

[4] SEforALL and Climate Policy Initiative (2019), Energizing Finance: Understanding the Landscape, Research Report

[5] Saeid Charani Shandiz, Greg Foliente, Behzad Rismanchi, Amanda Wachtel, Robert F. Jeffers (2020), *Resilience framework and metrics for energy master planning of communities*, Energy, Volume 203

[6] Hollnagel et al., (2006), *Resilience Engineering Concepts and Precepts*, Ashgate Edition

[7] M. Bazilian, P. Nussbaumer & al., (2011), *Energy Access Scenarios to 2030 for the Power Sector in Sub-Saharan Africa*, Fondazione Eni Enrico Mattei FEEM Working Paper No. 68.2011

[8] Anton Eberhard, Katharine Gratwick, Elvira Morella, Pedro (2017), *Independent Power Projects in Sub-Saharan Africa: Investment trends and policy lessons,* Energy Policy, Volume 108, Pages 390-424

[9] Sandia National Laboratories, U.S. Department of Energy (DOE) (2019). *Fundamentals of Advanced Microgrid Design*.

[10] United Nations Framework Convention on Climate Change. *Paris agreement on climate*. 2015. p. 1e16.

[11] Jesse B-J, Heinrichs HU, Kuckshinrichs W. *Adapting the theory of resilience to energy systems: a review and outlook*. Energy, Sustainability and Society 2019.

[12] Hay AH. *Surviving catastrophic events : stimulating community resilience*. 2013. p. 41-6.

[13] The World Bank. *Climate impacts on energy systems*. 2011.

[14] Rübbelke D, Vogele S. *Short-term distributional consequences of climate change impacts on the power sector: who gains and who loses?* Climatic Change 2013;116(2):191 - 206.

[15] Vugrin ED, Castillo A, Silva-Monroy C. *Resilience metrics for the electric power system: a performance-based approach.* Sandia National Laboratories; 2017.

[16] Bocchini P, Frangopol DM, Ummenhofer T, Zinke T. *Resilience and sustainability of civil infrastructure: toward a unified approach*. J Infrastruct Syst 2014;20(2):04014004.

[17] Andrew T. Jebb, Scott Parrigon and Sang Eun Woo, "*Exploratory data analysis as a foundation of inductive research*", Human Resource Management Review, vol. 27, no. 2, pp. 265-276, 2017, ISSN 1053-4822.

[18] Mario Li Vigni, Caterina Durante and Marina Cocchi, "*Chapter 3 - Exploratory Data Analysis Editor(s): Federico Marini Data Handling in Science and Technology*", Elsevier, vol. 28, pp. 55-126, 2013, ISSN 0922-3487, ISBN 9780444595287.

[19] A. S. Rao, B. V. Vardhan and H. Shaik, "*Role of Exploratory Data Analysis in Data Science*," 2021 6th International Conference on Communication and Electronics Systems (ICCES), Coimbatre, India, 2021, pp. 1457-1461, doi: 10.1109/ICCES51350.2021.9488986.

[20] X. Xu, H. Liu, M. Yao, (2019), *Recent Progress of Anomaly Detection*, Hindawi Complexity Volume 2019, article ID 2686378, https://doi.org/10.1155/2019/2686378

[21] J. Parmar, J. Patel,(2017), *Anomaly Detection in Data Mining: A Review, International Journal of Advanced Research in Computer Science and Software Engineering*, Volume 7, Issue 4, ISSN: 2277 128X

[22] S. S. Sahu, S. K. Pandey, and S. K. Singh, "*Anomaly detection in photovoltaic systems: A review*," Renew. Sust. Energy Rev., vol. 51, pp. 1219-1231, 2016.

[23] A. K. Singh and A. Kumar, "*Anomaly detection in photovoltaic systems using machine learning techniques: A review*," Renew. Energy, vol. 127, pp. 147-157, 2020.

[24] J. Liu, Y. Liu, and Y. Zhang, "*Anomaly detection in photovoltaic systems based on one-class SVM*," in 2018 IEEE 2nd International Conference on Control and Robotics Engineering (ICCRE), 2018, pp. 1-5.

[25] S. B. Dandin and S. R. K. Nair, "*Anomaly detection in photovoltaic systems using unsupervised machine learning algorithms*," in 2019 International Conference on Communication and Electronics Systems (ICCES), 2019, pp. 342-346.

[26] J. J. Lee, J. H. Kim, and S. H. Lee, "*Real-time anomaly detection in photovoltaic systems using convolutional neural networks*," in 2020 IEEE 11th International Conference on Advanced Technologies for Communications (ATC), 2020, pp. 1-5.

[27] He, Zhenyu, Xiaochen Zhang, Chao Liu, and Te Han. 2020. "*Fault Prognostics for Photovoltaic Inverter Based on Fast Clustering Algorithm and Gaussian Mixture Model*" Energies 13, no. 18: 4901. https://doi.org/10.3390/en13184901

[28] J. Keller, B. Kroposki, (2010), *Understanding Fault Characteristics of Inverter-Based Distributed Energy Resources*, NREL National Renewable Energy Laboratory US Dept. of Energy

[29] Rencher, Alvin C.; Christensen, William F. (2012), "*Chapter 10, Multivariate regression – Section 10.1, Introduction*", Methods of Multivariate Analysis, Wiley Series in Probability and Statistics, vol. 709 (3rd ed.), John Wiley & Sons, p. 19, ISBN 9781118391679.

[30] Efron, Bradley; Hastie, Trevor; Johnstone, Iain; Tibshirani, Robert (2004). "*Least Angle Regression*". The Annals of Statistics. 32 (2): 407–451. arXiv:math/0406456. doi:10.1214/009053604000000067. JSTOR 3448465. S2CID 204004121.

[31] Hooda, Nikhil & Azad, Amar Prakash & Panda, Pratyush & Saurav, Kumar & Arya, Vijay & Petra, M.. (2016). *PV Power Predictors for Condition Monitoring*. 10.1109/SmartGridComm.2016.7778763.

[32] Khan, Afaq Ahmad, et al. "*Fault Detection and Classification in Photovoltaic Systems Using Machine Learning Techniques: A Review*." Energies 14.1 (2021): 60.

[33] Wang, Tianyi, et al. "*A Review of Data-Driven Methods for Fault Detection and Diagnosis in Photovoltaic Systems*." IEEE Access 8 (2020): 160760-160772.

[34] Fathi, M., et al. "*Fault detection and classification of solar photovoltaic modules using machine learning algorithms: A comprehensive review*." Renewable and Sustainable Energy Reviews 134 (2021): 110391.

[35] Behnamfar, A., Nahman, J., & Huang, A. Q. (2019). *Review and challenges of fault detection and isolation techniques in microgrids*. Renewable and Sustainable Energy Reviews, 115, 109389.

[36] Deng, Z., Gao, W., & Liu, Y. (2016). *A novel adaptive threshold method for fault detection and isolation in microgrid.* IEEE Transactions on Power Systems, 31(1), 465-475.

[37] Gao, F., & Wen, J. (2015). *Review of non-linear control strategies for microgrid*. Renewable and Sustainable Energy Reviews, 45, 455-470.

[38] Hernandez, A., Cardenas, R., & Guerrero, J. M. (2017). *Review of fault protection strategies for single-and multi-microgrid systems*. Renewable and Sustainable Energy Reviews, 75, 154-170.

[39] Jiang, W., Li, G., Sun, H., & Wang, Y. (2014). *A novel transient fault detection method based on wavelet transform for microgrid*. IEEE Transactions on Smart Grid, 5(1), 150-158.

[40] Xu, Z., Zhu, J., Yang, S., Yang, J., & Xu, Z. (2015). *A hierarchical fault detection and isolation strategy for microgrid based on PCA and K-means clustering*. IEEE Transactions on Smart Grid, 6(2), 964-973.

[41] Wang, S., Fan, H., Tang, Z., Wang, H., & Wang, C. (2017). *A support vector machine approach for microgrid fault detection and classification.* IEEE Transactions on Industrial Electronics, 64(8), 6656-6666.

[42] Liu, H., Wu, Y., Dong, Z. Y., & Ren, Z. (2020). *A data-driven approach to microgrid fault diagnosis and prognosis.* IEEE Transactions on Industrial Electronics, 67(1), 451-461.

[43] Kathirgamanathan, A., Sriskanthan, N., & James, A. (2018). *A review of mini-grid systems in the developing countries: Perspectives, models, end-uses, economics, and policies.* Renewable and Sustainable Energy Reviews, 82, 1014-1030.