G0703

Managing and Optimizing a Set of PV Installations at the Low-Voltage Grid Level: A Data-Driven Concept through Machine Learning Techniques

Thibaud Alt (1), Beat Wolf (2), Jean-Philippe Bacher (3), Frédéric Montet (4)
(1) Groupe E SA, Route de Morat 135, 1763 Granges-Paccot/Switzerland;
Tel.: +41 26 352 52 52
thibaud.alt@groupe-e.ch
(2,4) iCoSys institute, HEIA-FR, HES-SO University of Applied Sciences and Arts, Western Switzerland, Bld de Pérolles 80, 1700 Fribourg/Switzerland;
beat.wolf@hefr.ch, frederic.montet@hefr.ch
(3) ENERGY institute, HEIA-FR, HES-SO University of Applied Sciences and Arts, Western Switzerland, Bld de Pérolles 80, 1700 Fribourg/Switzerland;
jean-philippe.bacher@hefr.ch

Abstract

This paper proposes a data-driven approach to managing and optimizing a set of photovoltaic (PV) installations by exploiting the possibilities of spatio-temporal modeling and machine learning techniques. Given the variable nature of solar energy production, optimizing PV installations for maximum output and efficiency is crucial. The aim is to identify trends, patterns, challenges, and opportunities for improvement in the operation of multi-site PV systems as well as to provide information for optimal management of the low-voltage network. A diverse array of methods are compared to forecast energy production, detect declines in system performance and refine maintenance scheduling. This study contributes to the growing field of renewable energy management by showcasing the effectiveness of ML models in optimizing a set of PV systems. It sets the stage for future progress in incorporating renewable energy sources into the electrical grid.

Introduction

The global energy transition, marked by a shift towards sustainability, highlights the urgent and significant challenges facing many countries in their current energy strategies. Among these, Switzerland, which aims to deactivate all its nuclear power plants by 2050 [1], has ambitious targets that require it to rapidly compensate for lost energy production and increase the integration of renewable energies. Against this backdrop, this paper proposes a data-driven approach to managing and optimizing a set of PV installations by exploiting the possibilities of spatio-temporal modeling and Machine Learning (ML) techniques. Given the fluctuating of solar energy production, optimizing PV installations for maximum output and efficiency is crucial. Integrating PV systems into the existing grid improves flexibility through self-consumption but brings with it a number of challenges related to grid stability and sizing. The aim is to identify trends, patterns, challenges, and opportunities for improvement in the operation of multi-site PV systems as well as to provide information for optimal management of the low-voltage network.
Using historical information on energy production, weather conditions, and PV systems specifications, solar production estimates derived from these models are compared with actual measurements to proactively detect underperforming installations. By analyzing this data, ML models are designed to map out standard operational patterns of PV plants, thereby identifying deviations indicative of potential issues. This methodology not only refines the accuracy of solar output forecasts but also offers an efficient way for the maintenance and optimization of PV installations, ensuring cost-effective and impactful interventions. A diverse array of methods ranging from conventional statistical methods, through Machine Learning (ML) algorithms, to advanced Deep Learning (DL) architectures are compared to forecast energy production, detect declines in system performance and refine maintenance scheduling. These methods were rigorously tested on a dataset, comprising multiple heterogeneous data from over 400 PV installations over a minimum historical period of two years, all situated within Western Switzerland.
This document is organized into several key sections. *Section 1: Context* provides an overview of the general context, outlines known issues, and reviews related works. *Section 2: Dataset* details the consolidated dataset that has been utilized. *Section 3: Methods* elucidates the techniques employed to identify declines in system performance. *Section 4: Results* showcases the findings from the study. *Section 5: Discussion* examines the results and proposes potential future directions. The final section, *Conclusion*, offers a comprehensive summary of the paper.

1. Context

Switzerland is currently facing several major challenges in its energy transition. The Federal Council has drawn up an energy strategy for 2050 [1], the main aim of which is to decommission the five existing nuclear power stations by then. In 2023 the total electricity consumed in Switzerland was 56.1TWh, marking a decrease of -1.7% compared to the previous year [2]. In the meantime, the Swiss have to find various ways to compensate for this 25 TWh production deficit, induced by the shutdown of these five power plants and the projected increase in consumption due to the decarbonization of mobility and housing sector. This transition requires coordinated measures and effective implementation to promote energy efficiency, increase the share of renewable energies, modernize electricity grids, and speed up administrative procedures. One of the building blocks of this transition is the use of renewable energies, and in particular the optimization and maintenance of existing PV installations. There are a number of areas where improvements can be made to maximize the contribution of existing PV installations to this energy transition. Firstly, it is essential to improve the performance of existing solar installations by using more advanced technologies and carrying out regular maintenance and cleaning of the solar

panels. This will ensure maximum solar electricity production. In addition, the integration of energy storage systems, such as batteries, means that the electricity generated can be stored for later use, guaranteeing continuous availability even when there is no sun. In addition, the integration of solar installations into smart grid systems offers more efficient management of the park and possible synergy options. Finally, raising awareness and training the owners and managers of solar installations in best maintenance and operating practices are crucial to maximizing their return.
The canton of Fribourg is aiming to significantly increase its production of PV solar energy by 2050 [3]. By September 2023, PV installations were already producing 200 GWh a year, and the aim is to reach 600 GWh by 2035 and 1,300 GWh by 2050. To achieve this, the canton is focusing on developing solar PV, making it an essential part of its future energy supply. Groupe E, the western Switzerland Distribution System Operators (DSO), connected the first PV installation to its grid twenty years ago. Since then, the pace has continuously accelerated, increasing from 500 installations per year a decade ago to over 5000 in the sole year of 2023. The number of connected PV systems has doubled since the end of 2020, reaching more than 20,000 installations by the end of 2023 [4]. Together, they represent a total installed capacity of 360 MW and an annual production of 360 GWh. In September 2023, Groupe E commissioned the largest solar farm in Switzerland, covering an area of 47,000 m2 including 19,000 panels. Operating this growing solar park presents a number of crucial challenges, particularly in terms of maintenance and servicing, performance monitoring and management, and grid integration. In addition, managing and exploiting the data generated by this fleet of installations, combined with in-depth analysis, offers the possibility of identifying trends and opportunities for improvement.
The main aim of this study is to make full use of the data produced by PV installations in order to optimize their contribution to solar energy production. This approach focuses on two main areas: firstly, the creation of a high-quality reference dataset to establish a solid and reusable basis for building models, and secondly, the implementation of a maintenance and in-depth analysis strategy.

2. Dataset

The first objective of this study is to provide a set of reference data. To achieve this, various data sources have been identified and collected. Thanks to the consolidation operations carried out on these data and the creation of a reference dataset, several PV installations are now available in the created dataset. This new dataset is structured in the form of tabular data made up of three tables presented in chapter 2.1. The total number of PV installations available for this analysis is 417, as shown in Figure 1.

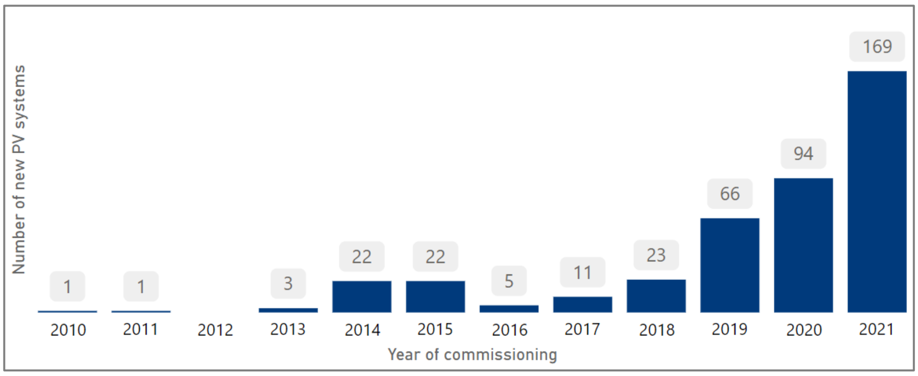

FIGURE 1 : NUMBER OF COLLECTED PV SYSTEMS DATA GROUPED BY COMMISSIONING YEAR

These installations are grouped by year of commissioning. Installations commissioned after 01.01.2022 are deliberately not included in the analysis. This decision is explained by the need to have consistent historical data to reliably assess changes in performance and characteristics over time. These 417 PV installations are spread across Western Switzerland in a geographically diverse manner, encompassing all the French-speaking regions and Bern. This balanced distribution ensures significant representation of diverse weather conditions and terrain. Each different environmental zone will allow a significant representation of specific climatic and topographical conditions. The geographical diversity of the sites means that potential variations in sunshine and altitude can be taken into account.
The various data sources collected and integrated into the reference dataset contain several dimensions. These dimensions have been divided into three tables detailed in the chapter 2.1. These tables will provide a structured framework for organizing and analyzing the data, facilitating a deeper understanding of the various dimensions and their interrelationships within the dataset.

2.1 Tables dimensions

The used dataset is structured in the form of tabular data made up of three tables: *pv_systems*, *telemetry* and *maintenance_cases*. Each table is interconnected via a distinct and shared unique identifier (UID), making it easier to identify the specific data associated with each PV site.

***pv_systems* —** This table is the heart of the reference dataset. It serves as a repository for a wealth of detailed information relating to PV systems. This information includes data on the installation, such as the date of installation, the type of installation and the technical specifications. In addition, precise geographical information allows each PV system to be located. The solar production capacity of each site is also documented, providing valuable insights into the energy potential of each installation. Information on the life cycle of each PV system is available, providing an overview of the sustainability and efficiency of these systems over time. The various components of PV systems, such as solar panels, inverters, and energy storage systems, are detailed. Finally, relevant administrative information, such as customer details, current system status and management information, is also stored in this main table.

***telemetry* —** The second table in the reference dataset encompasses a full spectrum of telemetry data. This table is a valuable resource that provides a detailed historical overview of energy production, consumption, and storage system status. Each record in this table is time-stamped, allowing rigorous chronological tracking of data for each installation. This makes it easy to analyze trends and patterns over time, providing valuable information for future system optimization. In addition, the table also incorporates meteorological data, including temperature, humidity, and wind speed. Together, these data form a comprehensive resource for the analysis and continuous improvement of energy performance.

***maintenance_cases* —** The third table in the reference dataset is a comprehensive compilation of recorded maintenance cases. This table includes details of the customer, the address of the installation, the specifics of the maintenance contract, the nature of the problem encountered, and the interventions carried out. Each entry in this table has time fields, enabling the duration of the interventions to be tracked. In addition, the table contains detailed information on the components replaced and the actions carried out during each intervention.

3. Methods

This section describes the diverse methods employed for several purposes: forecasting energy production, identifying decreases in system performance, and optimizing maintenance scheduling [5]. A range of techniques are explored and compared, from conventional statistical methods, through ML algorithms, to advanced DL architectures. By leveraging diverse methods and meticulously isolating system characteristics in different manners, the aim is to compare the results obtained through these techniques. It is conceivable that certain installations demonstrating insufficient performance across multiple techniques can be identified simultaneously. Subsequently, these identified underperforming installations be the subject of a more detailed analysis to pinpoint anomalies and facilitate predictive maintenance. For each method, the technique used, any ML models employed, and the data extracted from the reference dataset will be presented. The combination of results from different methods is presented in Chapter 4.

3.1 Statistical methods

This chapter presents the statistical techniques used, such as the calculation of the production index discrepancy analysis and the comparison of produced energy per installed surface.

3.1.1 Production Index Discrepancy Analysis

The first approach is a simple statistical method based on the use of a specific tool. For each PV installation, a production index is calculated using dedicated software. This index, which is fixed, is based on specific characteristics such as solar irradiation, the year's weather conditions, the PV technology used, the installed power, and the orientation and inclination of the panels. With each additional year of operation of the installation, the production index decreases by 0.8%, as shown by trending line on Figure 2. Then, for each installation, the delta between actual production and the index is calculated and consolidated as a percentage in a table. The facilities are then classified into 4 categories according to the following thresholds: above/corresponding to expectations (greater or equal to 100%), slightly below expectations (100% to 95%), below expectations (95% to 90%), and significantly below expectations (lower than 90%). The main drawback of this method is its assumption that the production index decreases by a fixed percentage with each additional year of operation.

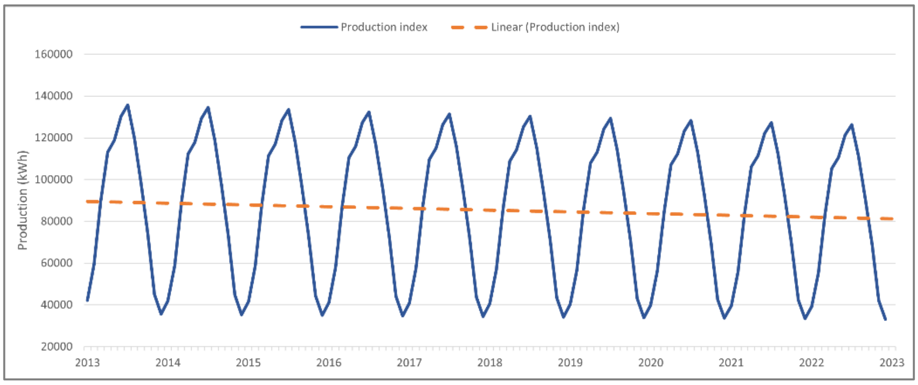

FIGURE 2 : ESTIMATED PRODUCTION INDEX OVER THE YEARS AND DECRESING TREND LINE

3.1.2 Produced energy per installed surface

This second method involves calculating the amount of energy produced by a PV system over a defined period normalized by the surface area of the PV panels. This normalization allows for a fair comparison between different installations, regardless of their size. The method initiates by calculating the cumulative energy output of each PV system throughout its operational lifespan, measured in kilowatt-hours (kWh). Next, the total energy production is divided by the surface area of the PV panels to obtain the energy produced per unit of the surface area, measured in square meters (m²). Once the energy production per surface area is calculated for each PV system, these values are compared with the average value calculated across all sites or installations under consideration. The performance of each PV system can be categorized into different groups or classifications according to the comparison results. An alternative approach would involve normalizing the data not based on the surface area, but rather by the stated peak power. However, this method was not selected due to the less reliable nature of peak power data. This approach has the advantage that the only necessary data are the sum of the energy produced over a given period of time and the installed surface area. One potential disadvantage of this method is that it may not fully account for variations in panel efficiency or environmental factors across different installations. Since it normalizes energy production solely by the surface area, it might overlook differences in panel quality, orientation, shading, or other factors that could affect performance. Additionally, it does not consider fluctuations in energy production due to factors such as weather conditions or maintenance issues, which could impact the accuracy of performance comparisons. Another disadvantage arises from the reliance on the average value calculated across all sites or installations in the dataset. If the dataset changes, this value will also change.

3.2 Clustering based methods

In this section, clustering, the first ML based method is introduced. Clustering techniques aim to partition a dataset into subsets. This is particularly useful for identifying inherent patterns and grouping data points based on similarity. In the context of energy production analysis, clustering algorithms can help uncover distinct clusters of PV installations with similar characteristics or production profiles [6]. Manual, k-means, and Density-Based Spatial Clustering of Applications with Noise (DBSCAN), the three clustering methods, focus solely on PV system specifications. These methods aim to group PV installations according to inherent similarities in their characteristics. The produced energy is normalized by the surface area to account for differences in the size of the PV panels. This normalization process reduces the available PV systems in the main dataset from 417 to 99 for which information on the surface area is available. Once the PV systems are distributed into clusters, the analysis continues by calculating the production average for each group. This average represents the typical production of installations within that particular group. Subsequently, each installation is compared to the production average of its corresponding group. This comparison allows for the assessment of each installation's relative performance compared to its peers. A delta is calculated by subtracting the actual production of each installation from the production average of its group. This delta represents the difference between the individual performance of each installation and the average performance of its group.

3.2.1 Manual clustering

This approach involves the manual categorization of PV systems based on specific attributes. While it lacks the automation of algorithmic clustering, manual clustering allows for expert domain knowledge to guide the grouping process, potentially uncovering nuanced patterns not captured by automated methods. The features selected for manual clustering are main panels orientation and main panels inclination. Only two features have been chosen for manual clustering because they are deemed to have the most significant impact. This configuration results in the formation of 15 clusters.

3.2.2 k-means clustering

The k-means algorithm is a popular clustering method that partitions a dataset into a predetermined number of clusters by minimizing the sum of squared distances within each cluster. In this study, k-means was used to group installations with to groups of features.

The first set (A) includes: main orientation degree, main inclination, latitude, longitude, altitude, category, and type. The second set (B) incorporates the features from the first set as well as the roof type, the roof slope, the panel manufacturer, and the inverter manufacturer. The optimal result with k-means is achieved using the second set (B) with 29 clusters.

3.2.3 DBSCAN clustering

Unlike k-means, DBSCAN does not require specifying the number of clusters beforehand. Instead, it identifies clusters based on density connectivity, grouping together data points that are closely packed while labeling outliers as noise. DBSCAN is particularly useful for identifying irregularly shaped clusters and handling noise effectively, making it suitable for clustering PV installations with varying densities and spatial distributions. In this study, DBSCAN was used to group installations with the same feature sets as k-means (sets A & B). Once clustered, each site is labeled with a specific group number. By comparing the members of each group, it is possible to identify potential correlations between the different characteristics of PV systems, thereby gaining a better understanding of their respective performances. The optimal result with DBSCAN is achieved using the second set (B) with 10 clusters.

3.3 Produced energy predictions

In this section, the first DL techniques are introduced. The core aim of this approach is to harness the power of ML and DL methods to generate precise forecasts of energy production for individual PV systems on a daily basis. By using ML and DL techniques to forecast the energy production of individual PV systems, and then comparing this with the actual output, it is then possible to obtain a performance score. Three distinct methods have been developed to anticipate energy production. The first, detailed in chapter 3.3.1, relies only on the analysis of historical telemetry data collected from each PV installation. It aims to predict daily time series of production. The second approach, detailed in chapter 3.3.2, integrates specific installation characteristics, in addition to incorporating aggregated data on annual production. The third and final method, detailed in chapter 3.3.3, uses only a few of the technical characteristics of PV installations and therefore allows a prediction based on a restricted set of data. The second and third methods aim to predict energy production on an annual basis.

3.3.1 DeepAR model

For this prediction approach, production data is collected and summarized by daily produced energy (kWh). With this daily basis, the *DeepAREstimator* class from the *GluonTS* framework is used to train a DeepAR prediction model. *GluonTS* is a Python package for probabilistic time series modeling, focusing on DL-based models, that includes *DeepAR* model. The DeepAR model is a supervised learning algorithm specified for forecasting time series by using Recurrent Neural Networks (RNN) [7]. Once all predictions are completed, the results are aggregated for each PV system by week, and the delta with the real production of the year is calculated. This delta provides a performance index as detailed in Chapter 4.

3.3.2 Fast and Lightweight AutoML (FLAML)

For this second prediction approach, the FLAML Python library has been chosen. FLAML library can find accurate ML models automatically, efficiently, and economically from a defined set of settings [8]. The used set of features used to predict the produced energy includes installation year, surface, main orientation degree, main inclination, latitude, longitude, altitude, category, type, roof type, roof slope, panel power and the total energy produced for the years 2018 to 2022, if the data are available. After multiple training sessions, the best-found model is an XGBoost with 13 estimators, which provides a good balance between performance and efficiency. These estimators correspond to the number of trees boosted by the gradient. Once the best-performing model has been trained, predictions are made on the test set, and the delta between the predicted and actual values is calculated.

3.3.3 XGBoost with limited depth model

This third prediction approach has the advantage of using only a small fraction of the characteristics of PV systems. These characteristics are the date of commissioning, the peak power, the orientation and inclination of the panels, the type (residential or commercial) and a geographical area determined by the location of the PV system. Several ML models were tested, including Light Gradient-Boosting Machine, Random Forests, XGBoost and CatBoost [9]. The most promising results were achieved with the XGBoost with limited depth model, employing 10 estimators and a maximum depth of 6.

4. Results

In this chapter, the results of applying the various methods outlined in Chapter 3 are showcased, and a comparative analysis of the outcomes yielded by these techniques is conducted. The goal here is to compare the results of the various methods used and to identify the PV systems that show lower performance according to several of these methods. Thanks to the Production Index Discrepancy Analysis method, five installations with problematic production for at least three years were identified. These installations, referred as witnesses, can be used as a baseline for other methods.

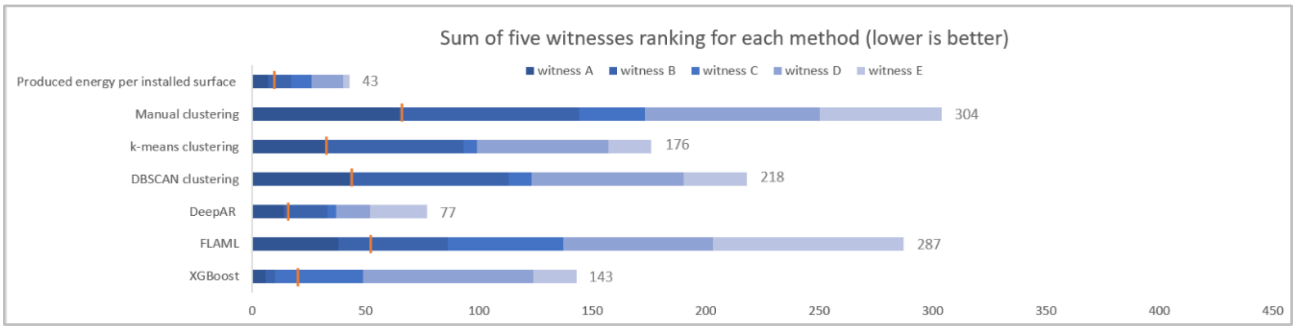

FIGURE 3: THE SUM OF RANKING OF EACH WITNESS BY METHOD, THE BESTVALUE IS THE LOWEST

As shown on Figure 3, methods are compared by the total ranking of the five witnesses. The scale ranges from 15 (the baseline score) to 485 (the worst score). The number in gray indicates the sum of the ranking, while the orange line shows the median. Produced energy per installed surface and DeepAR methods rank below 100. The rank of the highest witness for produced energy per installed surface method is 14, while the median rank is 9. For the DeepAR method, the highest ranking is 25, while the median rank is 15. Clustering-based methods achieved the highest rankings of 79 for manual, 61 for k-means, and 70 for DBSCAN, with median rankings of 65 for manual, 32 for k-means, and 43 for DBSCAN. FLAML method obtains the highest ranking of 84 and a median of 51. Finally, XGBoost with limited depth method obtains a last witness ranked at position 75 and a median ranking of 19.

On Figure 4, the performance breakdown of PV systems is grouped into 25% intervals. The number of for each interval is displayed in white, while the red line indicates the baseline of well-functioning systems. Systems that underperform are those that repeatedly appear in intervals below 75%.

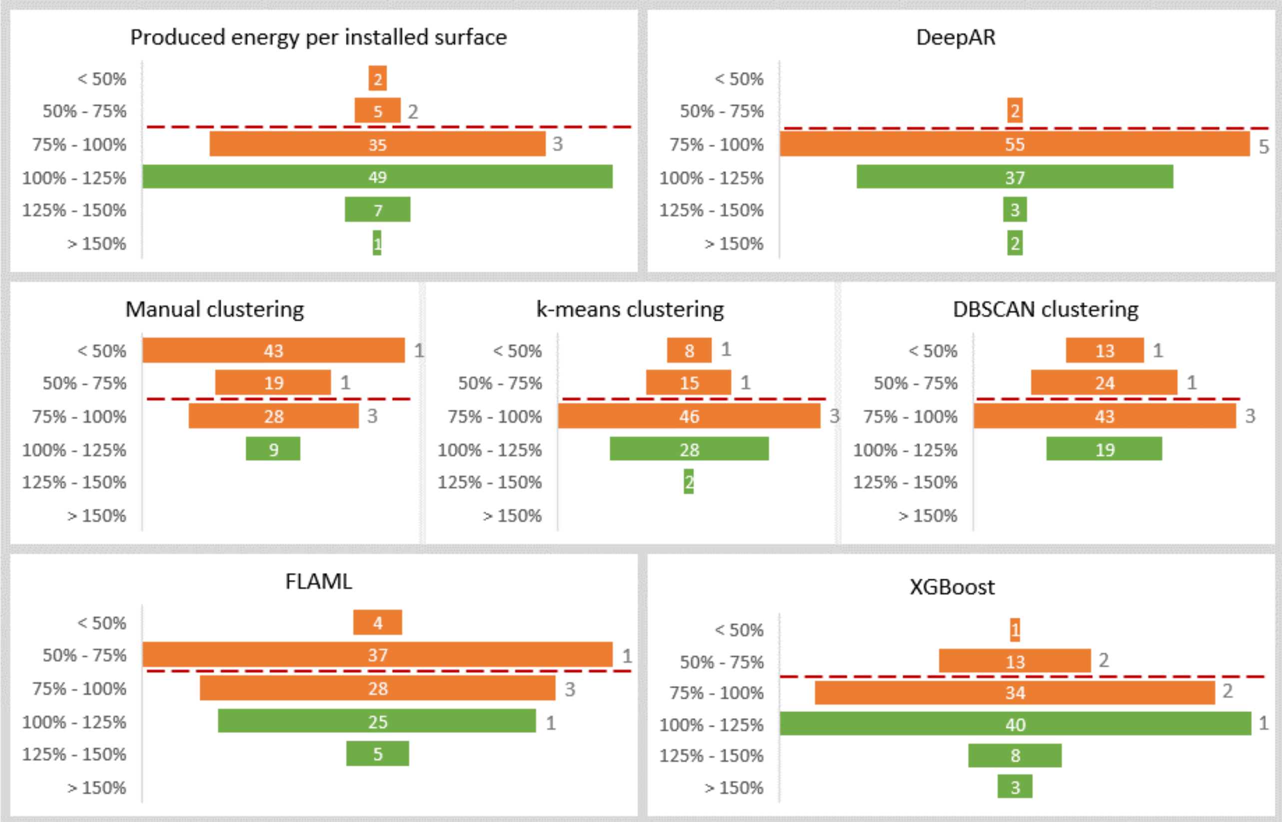

FIGURE 4: COMPARISON OF THE PERFORMANCE BREAKDOWN OF PV SYSTEMS FOR EACH METHOD, NUMBER OF WITNESSES IN GRAY

Produced energy per installed surface method identifies a total of 7 underperforming PV systems, including 2 of the 5 benchmark systems. The tree clustering-based methods categorize 1 PV system in the first interval, 1 in the second interval, and 3 in the 75%-100% interval. Manual clustering indicates that 62 PV systems are underperforming, while k-means identifies 23 underperforming systems, and DBSCAN identifies 37. Surprisingly, DeepAR model includes more than half of the dataset within the 75%-100% interval, including the 5 witnesses. Only two systems are considered significantly underperforming. Lastly, the FLAML and the XGBoost methods are the only ones that classified witnesses in an interval greater than 100%, which suggests that the results may not be reliable. However, these methods still indicate 41 and respectively 14 under-performing PV systems.

5. Discussion

The *produced energy per installed surface method* achieves the best results in ranking the witnesses. However, it requires precise production data and a significant common denominator, such as surface area or peak power. Unfortunately, these essential elements are unavailable for more than 75% of actual PV systems. *Clustering based methods* achieve mixed results in ranking the witnesses. However, their primary advantage is that they do not require production information and are only based on technical characteristics. The accuracy of their ranking could be enhanced by carefully selecting features that are essential and available on most of the PV systems currently in service. The *DeepAR model*, like statistical methods, requires production data, which shares the same disadvantages. However, this data can be aggregated, requiring less precision. It achieves accurate results for the ranking of witnesses. It obtains accurate results for the ranking of witnesses. The *FLAML method* has the advantage of only requiring some technical data and an annual estimation of the produced energy, but it falls short in terms of witness detection. The *XGBoost with limited depth model* produces encouraging results and has

the immense advantage of requiring a limited number of features. All methods have a recall of 1 (every witness is included in intervals below 100%), except for the FLAML and the XGBoost methods, which have a recall of 0.8.
With the results established in Chapter 4, and considering only results below 75% from each method (excluding witnesses), it is observed that 12 PV systems are deemed underperforming by four or more methods. Among these, ten installations are considered poorly performing according to four methods, one according to five, and another according to six. This discovery is a significant outcome of the study, signaling potential areas for intervention and optimization. However, formally confirming the status of these installations would require on-site verification, introducing logistical challenges. Future research directions would focus on detecting anomalies in specific sites that have been identified as abnormal or underperforming, starting with the first group identified. Additionally, the aim is to explore other methods as presented in Chapter 3 like a fourth clustering method, called Time Series-Based Clustering, leveraging production time series data [10]. Another way would be to use advanced DL architectures like Graph Neural Networks (GNNs) and Transformers [11], Transfer Learning (TL) [12], or zero-shot forecasting [13] to enhance detection of PV systems that show abnormal performance without needing the energy data produced.

Conclusion

In this paper, an innovative data-driven approach to find underperforming PV installations is proposed. At a time when the global energy transition demands increased integration of renewable energies, the proposed approach aims to leverage historical data of energy production, weather conditions, and PV system specifications to optimize their performance and efficiency. By analyzing this data, a comparison of a range of methods has been conducted, spanning from conventional statistical methods to advanced clustering and DL techniques. There is no single method: the choice depends on the data available. Depending on the data available, one method or another may be more appropriate. The results show that several installations are performing below expectations and require special attention. However, it is acknowledged that challenges persist, including on-site verification of identified underperforming installations and more specific anomaly detections.
In conclusion, this study marks a significant milestone in the management and optimization of PV installations at the low-voltage grid level, providing data-driven solutions to address the challenges of the energy transition towards a more sustainable future. The findings emphasize the importance of continuous innovation and adaptation in renewable energy management, setting the stage for further advancements in incorporating renewable energy sources into the electrical grid. Future research will focus on enhancing clustering methods and integrating more sophisticated ML models to achieve even greater accuracy and reliability in PV system performance optimization.

References

[1] T. E. a. C. Federal Department of the Environment, “Principles of energy policy,” [Online]. Available: https://www.uvek.admin.ch/uvek/en/home/energy/energy-strategy-2050.html. [Accessed 02 05 2024].

[2] S. F. O. o. Energy, “Electricity consumption down by 1.7% in 2023,” 18 04 2024. [Online]. Available: https://www.admin.ch/gov/en/start/documentation/media-releases.msg-id-100748.html. [Accessed 02 05 2024].

[3] S. d. I. d. I. d. Fribourg, “Le Conseil d’Etat adopte une stratégie solaire photovoltaïque ambitieuse,” 01 09 2023. [Online]. Available: https://www.fr.ch/deef/sde/actualites/le-conseil-detat-adopte-une-strategie-solaire-photovoltaique-ambitieuse. [Accessed 02 05 2024].

[4] G. E. SA, “Le réseau électrique de Groupe E a accueilli sa 20’000e installation photovoltaïque,” 21 12 2023. [Online]. Available: https://www.groupe-e.ch/fr/decouvrir-groupe-e/medias/communiques-de-presse/20000-installations-photovoltaiques. [Accessed 02 05 2024].

[5] U. K. Das, K. S. Tey, M. Seyedmahmoudian, S. Mekhilef and M. Y. I. Idris, “Forecasting of photovoltaic power generation and model optimization: A review,” *Renewable and Sustainable Energy Reviews,* 01 2018.

[6] Z. a. K. I. a. R. M. Wang, “Clustering Based Methods for Solar Power Forecasting,” *International Joint Conference on Neural Networks (IJCNN)At,* 07 2016.

[7] D. Salinas, V. Flunkert, J. Gasthaus and T. Januschowski, “DeepAR: Probabilistic forecasting with autoregressive recurrent networks,” *International Journal of Forecasting,* 07 2020.

[8] D. P. Kumar, N. Kaur and S. S. Kushana, “Efficient Time Series Photovoltaic Production Forecasting using FLAML with Traditional Machine Learning Algorithms,” *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS),* 17 05 2023.

[9] D. Adhya, S. Chatterjee and A. K. Chakraborty, “Performance assessment of selective machine learning techniques for improved PV array fault diagnosis,” *Sustainable Energy, Grids and Networks,* 03 2022.

[10] Z. Zhang, J. Wang, Y. Xia, D. Wei and N. Yunbo, “Solar-Mixer: An Efficient End-to-End Model for Long-Sequence Photovoltaic Power Generation Time Series Forecasting,” *IEEE Transactions on Sustainable Energy,* 10 2023.

[11] J. Simeunović, B. Schubnel, P.-J. Alet, R. E. Carrillo and P. Frossard, “Interpretable temporal-spatial graph attention network for multi-site PV power forecasting,” *Applied Energy,* 12 2022.

[12] S. M. Miraftabzadeh, C. G. Colombo, M. Longo and F. Foiadelli, “A Day-Ahead Photovoltaic Power Prediction via Transfer Learning and Deep Neural Networks,” *Forecasting,* 17 02 2023.

[13] V. Ekambaram, A. Jati, N. H. Nguyen, P. Dayama, C. Reddy, W. M. Gifford and J. Kalagnanam, “Tiny Time Mixers (TTMs): Fast Pre-trained Models for Enhanced Zero/Few-Shot Forecasting of Multivariate Time Series,” 09 04 2024.