Towards Data-Driven Malfunctioning Detection in Public and Industrial Power Grids

David Fellner*, Helfried Brunner*, Thomas I. Strasser*†, Wolfgang Kastner†

†AIT Austrian Institute of Technology, Vienna, Austria {david.fellner, helfried.brunner, thomas.strasser}@ait.ac.at
‡TU Wien, Vienna, Austria {thomas.strasser, wolfgang.kastner}@tuwien.ac.at

***Abstract*—Investigation of a concept for remote detection of malfunctioning grid supporting devices using minimal data is proposed in this work. The operation of future electricity grids highly depends on the behaviour of these devices and their support functions such as reactive power dispatch used for voltage control. Synthesising and utilizing operational data of a distribution grid, a functionality is being developed to enable better surveillance of grid connected devices ensuring security of supply and resiliency. In a first step, data driven approaches for anomaly detection are explored. They are applied to the operational data of the device to detect unwanted behaviour. Results show first indicators of applicability and possible obstacles.**

I. Introduction

Nowadays, power grid operators face many challenges connected to the fundamental changes the energy system is undergoing. However, this transformation is essential in the shift to a new green and sustainable energy system. The decentralization of power generation causes many of these problems [1]. These include regulatory issues, environmental concerns as well as technological challenges, such as transmission and storage of electrical energy. Especially a high density of photovoltaic (PV) power generation has big influence on grid operation [2]: PV generation exceeding local energy demand leads to reverse power flows from lower voltage levels to the transmission system as well as voltage rises.

Globally this can impact the system frequency. Locally violations of the admissible voltage magnitude, the so called voltage band, are a common consequence. To limit renewable energy generation as little as possible but avoid such unfavourable effects at the same time, control strategies are needed. To integrate distributed generation in distribution networks, voltage regulation is regarded as the principle measure [3]. This is implemented through grid supporting functionalities (frequency, voltage, etc.) provided by the generation units. These include curtailing the active power dispatched as well as controlling the reactive power dispatch of generation units with inverters, most commonly via a local droop control [4]. The functions of these controls are shown in Figure 1: as depicted on the left, the power factor, and therefore also the reactive power is controlled depending on the active power. The right side of Figure 1 depicts a voltage control varying the reactive power. Generation units are obliged to exercise certain control mechanisms as stated in national regulations [5]. A reactive power control depending on the local voltage has to be implemented even by the generation units with the least rated power listed. Furthermore, the proper amount of reactive power actually has to be dispatched.

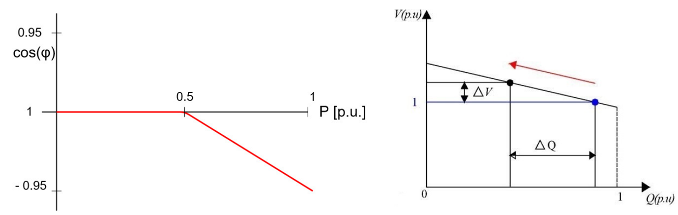

Fig. 1: Q(P) (left) and voltage droop control (right) schemes [6].

A data driven approach for monitoring these functions is recommendable as it keeps the required knowledge about network components characteristics to a minimum. Grid components data is in many cases not updated regularly or often faulty. Therefore, an approach using operational data is of advantage [7]. However, this data is not always available at all connection points. This can be due to the lack of measurements or legal restrictions on usage of these for reasons of data privacy. Means of surveillance ought yet to be developed to detect and distinguish transformer malfunctions or wrongly paramaterized PV inverters under these circumstances. In this paper, a first approach on detecting the latter is described. This paper also gives a first outlook on which measurements, such as voltage, are best suited for this task, as well as how high a quality of these measurements is necessary.

The paper is organized as follows: Section II outlines the state-of-the-art in the domain, targeted scenarios are introduced in Section III. An overview of the methodological approach is presented in Section IV, whereas preliminary results are discussed in Section V. Finally, a conclusion and an outlook on the work-in-progress is given in Section VI.

II. Related Work

To ensure the intended behaviour is actually being executed by devices, anomalies during operation have to be detected. Therefore, methods to detect these anomalies are needed. Nevertheless, no framework covering the entirety of this task can be found in literature. Yet, many approaches exist presenting solutions to parts of the challenges raised.

In [8], a number of algorithms are implemented and compared, including k-means or fuzzy c-means. These are applied on data of medium voltage transformers at main substations to cluster their consumption patterns. Followingly, to identify unusual consumptions in hourly load data, the local outlier factor (LOF) is used. Certain traits of the consumption such as irregular peak unusual consumption, broadest peak demand, sudden large gain and nearly zero demand unusual consumption are employed to characterize unusual patterns. Whether the characteristics utilized for detection can be applied on a wider spectrum of problems is yet to be determined.

The research in [9] presents a combined model of Gaussian distribution and polynomial regression. Using these, anomalies are detected in the modelled electrical energy consumption data of various schools. For anomaly detection on a household or grid connected device level, this approach is well suited. Nevertheless, it will most likely be necessary to create models for each application automatically, which is not included in this approach.

Extraction of typical consumption profiles of buildings is presented in [10]. Classification of buildings into categories s done beforehand. These reflect several characteristics such as type, physical properties or environmental conditions, for example the weather. For each class, typical load profiles are modelled based on these. The authors indicate the possibility to apply the method presented on fault detection and diagnosis (FDD). However, a discussion about the details of implementation of such an application is missing.

Summarising, an approach for anomaly detection and classification tailored to the power distribution grid is yet to be explored. Therefore, the particular challenges of identifying models automatically as well as the applicability of algorithms to the data at hand have to be addressed.

III. Targeted Scenarios

Both a solution applicable to public as well as to industrial distribution grids is desirable. Therefore, scenarios are to be explored taking into account the peculiarities of both: in industrial environments distribution distances are rather short, power densities are frequently high. For these reasons the grid experiences heavy loading. Furthermore, regulations applying to public power grids, such as ones concerning data protection, might not be relevant for industrial grids [11]. The prevalence of big, fast varying loads in industrial grids such as excavators also leaves an impact on protection schemes as mentioned in [12]. This makes the detection of malfunctions particularly interesting here. Moreover, surveillance of the correct scheduling of such loads in a context of factory automation for production planning with regard to demand side management can potentially be a rewarding use case [13]. In general only detection of non dynamic operational changes are addressed. Therefore, only data in minute resolution or lower is going to be necessary and used. The cases targeted include the surveillance of the correct switching of loads or the proper dispatch of energy by generation units, which are all non transient events.

IV. Methodological Approach

Data driven approaches as machine learning algorithms should be implemented and evaluated. Furthermore, the data used has to be synthesised.

A. Anomaly Detection

Kernel principle component analysis (kPCA) [14] seems to be a fitting solution for anomaly detection. As presented in [15], it can be used to build a statistical model of a systems nominal state. In cases where several anomalous states are possible but only little information is accessible about them, but a stable nominal state is usually encountered, kPCA is especially well suited. Foremost because it is unsupervised. For both grid participants at the low voltage level as well a transformer this can be expected. An alternative method incorporating information from labeled, unlabeled and partially labeled data is posed by the partially hidden structured support vector machine (pSVM) based on [16]. This method can help by revealing hidden structures in the data. Additionally, relations between 'stable' and 'different' events can be captures using [15]. Furthermore, [17] explores a one class support vector machine for anomaly detection in Heating, Ventilation and Air Conditioning (HVAC) systems. It's pointed out, that abnormal events do not cause the biggest variance in the data, but the regular operation of the system does. This is also applicable to households or PV generators showing usual functionalities. Therefore, on the low voltage level the usage of primary component analysis (PCA) for anomaly detection appears advisable. On the medium voltage level applying a support vector machine (SVM) seems promising.

B. Software, Data Acquisition and Validation

Grid simulation is to be used to synthesize data of scenarios in which devices that usually provide grid support functionalities experience malfunctions. These malfunctions cause them to stop providing this grid support. This ought to be detected by the impact this leaves on grid operational data. The grid participants should therefore act as life like as possible. This can be aided by applying real world load curves of household or industrial loads and energy dispatch patterns of generation units, such as PVs, to them. Data of regular operation and abnormal behaviour, such as in the case of load switching or incorrectly parameterized control patterns, are needed. Data of distribution transformers as well as low voltage grid participants data comprising of voltages, currents, and power flows should be generated in this manner. The low voltage data should be synthesised at the devices in a way to mimic smart meter measurements. This can be achieved by adding noise to the generated, clean data. The data acquired by these steps should then be used to develop, test and improve the anomality detection approaches presented above. Misclassification rate or a confusion matrix representing false negatives and false positives of the anomaly detection can function as key performance indicators (KPIs), helping to evaluate the applicability of the different approaches. As false alarms ought to be avoided, false positives are of particular interest here. Ideally, DSOs or operators of industrial grids could provide real world data to be used to develop the concept further and verify it. Furthermore, to test the method, the approach could be implemented at a grid serving as a demo site.

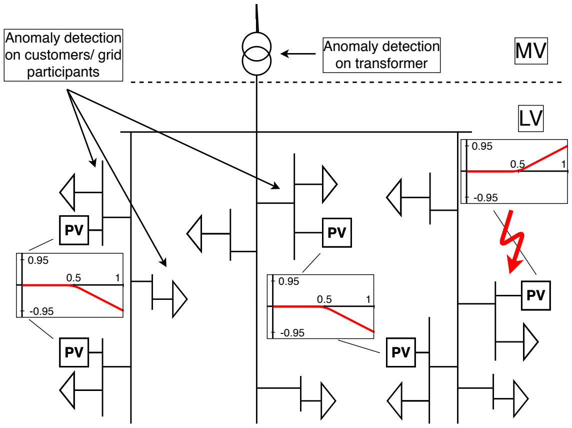

Fig. 2: Grid model used to generate data.

V. Preliminary Results

The exploration of a functionality that is able to monitor grid supporting devices is currently undertaken. Therefore, a framework to synthesise the necessary data was implemented using a grid modelling software (DIgSILENT PowerFactory) as a solver controlled by a Python API. A typical low voltage (LV) distribution grid consisting of 125 load terminals on 6 feeders, 3 of them in a radial configuration, the other 3 operated in an open ring topology, was modelled for this purpose. The 0.4kV LV grid is connected to the 20kV medium voltage level by a single 630kVA Dyn-connected distribution transformer. Every terminal has a household load connected to it, but only a certain share of the households are equipped with PV. The terminals to which a PV is connected are chosen randomly at the beginning of the simulation. The topological setup is in a symbolic manner depicted in Figure 2. All loads and PVs are connected to a terminal, which is in turn connected to a feeder via a line. In the simulations, the data is recorded at these terminals, which would represent smart meters in the real world, as well as on the transformer terminal. To mimic actual smart meter measurements, noise is added to the data created. The smart meters assumed are standard smart meters with an accuracy of 1%, meaning the error on the reading is at max 1% of the smart meters highest displayable value. Therefore, a Gaussian white noise, with a mean of 0 and a standard deviation such that the distributions biggest value is about the maximum error of the smart meter, is added to the data.

Each household and PV follows real consumption or generation profiles, respectively. The household profiles are available at a 1 minute time resolution, whereas the PV profiles have a time resolution of 5 minutes. 9 different PV profiles as well as 14 household profiles are distributed randomly to the loads and generation units. This creates various combinations of load and generation patterns at terminals. The household load profiles are scaled in a manner as to leave the transformer with a maximum loading of 70% over the course of a year. The PV profiles are scaled accordingly. The scaling also considers the strength of a connection point in the form of its hosting capacity normalized to the average of all terminals. The hosting capacity here is the maximum real power drawn before a voltage band limit is reached. All PV units are parameterized to follow the same control curve regarding reactive power dispatch. The curve applied is a cosϕ(P) as shown on the left side of Figure 1, only with a maximum power factor of 0.9 as required by German national regulations [18].

In the scenario simulated, the share of terminals to which a PV is connected was chosen as about 25%, equipping 33 terminals with PV generation. The simulation is run over the course of an entire year, with a simulation step size of 5 minutes. For every simulation run, the positions of the PVs, the PV inverter device which is experiencing a malfunction as well as when this malfunction occurs are chosen randomly.

The malfunction assumed in this scenario is a single PV inverter resetting itself to its default settings for unknown reasons. This deactivates any control curves parameterized. In this case, this function is a reactive power dispatch curve controlled by the active power, which influences the voltage at the terminal the PV is connected to. This means the device is failing to inject reactive power in times of high active power dispatch meaning not exercising its grid support function. From a certain point in time on, the voltage is therefore not controlled accordingly.

This malfunction is to be detected by analysing its impact on the grid. Figures 3 and 4 depict a first approach on this and its preliminary result: all figures show voltages at terminals with both households and PVs connected to them. Every graph shows the voltage at two of these terminals plotted against each other. Each data points x component is the voltage of the one terminal, the y component is the voltage at the terminal at a certain point during the simulation. A time window of 48 hours is chosen, meaning the graphs depict data starting 24 hours before the simulated malfunction and ending 24 hours after the occurrence of the malfunction.

Figure 3 shows the raw generated data. Figure 4 presents data with the noise described above added. The graphs on the right side of both figures depict voltages of two terminals plotted against each other where the PV inverters show regular behaviour. Especially in Figure 3 it appears legitimate to assume that voltages at terminals with correctly functioning devices are highly correlated, easily understandable from the datas appearance as a line. In contrast to this, the left hand side of Figure 3 shows voltages at two terminals where one of them has a PV generator connected that experiences the before mentioned malfunction. The point cloud does obviously not resemble a line quite as much in this case. This behaviour offers the possibility to extract the point clouds primary components as features for anomaly detection. In all graphs the red arrow indicates the first primary component, the one reflecting the biggest variance in the data. The green arrow points in the direction of the second biggest variance of the data, and is therefore the second primary component. Now these primary components explain certain shares of variance. In the case of the two terminals without malfunctions this share is 0.04%. For the terminals of which one experiences a malfunction, the second primary component explains 1.64%

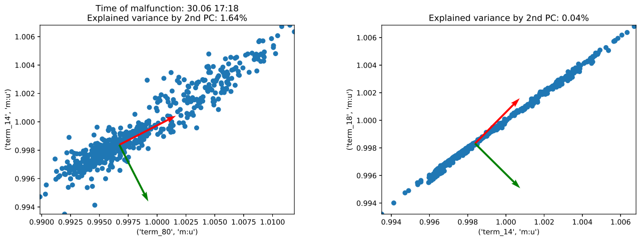

Fig. 3: Voltages plotted against each other.

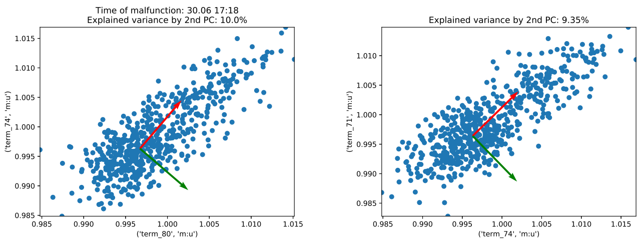

Fig. 4: Noisy voltages plotted against each other.

of the variance. Therefore, a change in this share could be used to detect anomalies. As striking as this discrepancy is in the raw synthesised data without added noise, the picture becomes a whole lot more blurry when noise is added to the data. In Figure 4 the shares of variance explained by the second primary components have risen to 10% for the malfunctioning case on the left and 9.35% for the terminals showing regular behaviour on the right. The diminished, yet still existent gap, in this share indicates that it might be harder to implement a detection in this way, but even with noisy data it might still be possible. This illustrates how important an evaluation of necessary and available data quality is.

VI. Conclusions and Outlook

The inevitable integration of decentralized renewable energy sources necessary to ensure a sustainable electric power supply rises challenges. Problems in grid operation can be caused by the great volatility these generation units show when providing energy. Therefore, these units have to provide grid support functionalities and grid operators have to be able to ensure that the same are exercised correctly. For that reason the goal of this work is a validated method allowing central and remote surveillance of expected functions. This is done via detection of changes in the behaviour of grid connected devices. These include generation units and their inverters.

The first interim results appear to be promising. Detection of malfunctions based on operational data, such as the voltage, appears to be an applicable solution, especially if clean data is available. Detection seems to be a lot more challenging on noisy data. This underlines the importance of reproducing real world circumstances accurately. Further investigation regarding the choice of data as in which terminals to compare or how big a time window is used for detection is still necessary.

Further work will also include classification of anomalies detected which should help to categorize the unwanted behaviour and refer to a cause. This should be possible on the medium voltage level using operational data at a transformer, as well as on the low voltage level. Therefore, the transformer load profile could be disaggregated into its contributions for data mining. The approach explored should only use relevant data for this purpose in order to limit data traffic and avoid legal issues. The framework will also be tested with some selected use cases in order to prove its usability. The use cases will consist of simulation examples, but also real world data to test the tool would be desirable. Concepts and methods on how to develop such a tool are also expected as further interim results.

Acknowledgment

This work received funding from the Austrian Research Promotion Agency (FFG) under the “Research Partnerships – Industrial PhD Program” in the DeMaDs project (FFG No. 879017). Furthermore, developments related to this work are involved with the ”PoSyCo” project (FFG No. 867276).

References

[1] E. Brown *et al.*, “Governance, decentralisation and energy: a critical review of the key issues,” 2015.
[2] J. von Appen *et al.*, “Time in the sun: The challenge of high pv penetration in the german electric grid,” *IEEE Power and Energy Magazine*, vol. 11, no. 2, pp. 55–64, March 2013.
[3] N. Mahmud and A. Zahedi, “Review of control strategies for voltage regulation of the smart distribution network with high penetration of renewable distributed generation,” *Renewable and Sustainable Energy Reviews*, vol. 64, pp. 582 – 595, 2016.
[4] V. Barrios *et al.*, “Feasibility and performance assessment of commercial pv inverters operating with droop control for providing voltage supports services,” in *ISGT Europe*, 2019.
[5] *Technische und organisatorische Regeln für Betreiber und Benutzer von Netzen*, E-Control Std., 2019.
[6] S. Conti *et al.*, “Optimal dispatching of distributed generators and storage systems for mv islanded microgrids,” *IEEE Trans. Power Delivery*, vol. 27, pp. 1243–1251, 07 2012.
[7] W. Luan *et al.*, “Smart meter data analytics for distribution network connectivity verification,” *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1964–1971, July 2015.
[8] D. D. SHARMA *et al.*, “Identification and characterization of irregular consumptions of load data,” *Journal of Modern Power Systems and Clean Energy*, vol. 5, no. 3, pp. 465–477, May 2017.
[9] W. Cui and H. Wang, “A new anomaly detection system for school electricity consumption data,” *Information*, vol. 8, p. 151, 11 2017.
[10] A. Capozzoli *et al.*, “Mining typical load profiles in buildings to support energy management in the smart city context,” *Energy Procedia*, vol. 134, pp. 865–874, 10 2017.
[11] M. Ingo and K. Rajaniemi, “Industrial smart grid,” Tech. Rep., 2013. [Online]. Available: https://bit.ly/3gOif8e
[12] M. A. Haj-ahmed and M. S. Illindala, “Investigation of protection schemes for flexible distribution of energy and storage resources in an industrial microgrid,” in *I&CPS*, 2014, pp. 1–10.
[13] M. Paulus and F. Borggrefe, “The potential of demand-side management in energy-intensive industries for electricity markets in germany,” *Applied Energy*, vol. 88, no. 2, pp. 432 – 441, 2011.
[14] H. Hoffmann, “Kernel pca for novelty detection,” *Pattern Recognition*, vol. 40, pp. 863–874, 03 2007.
[15] Y. Zhou *et al.*, “Abnormal event detection with high resolution micropmu data,” in *PSCC*, June 2016, pp. 1–7.
[16] Y. Zhou *et al.*, “Veto-consensus multiple kernel learning,” in *Thirtieth AAAI Conf. on Artificial Intelligence*, 02 2016.
[17] A. Beghi *et al.*, “A one-class svm based tool for machine learning novelty detection in hvac chiller systems,” *IFAC Proceedings Volumes (IFAC-PapersOnline)*, vol. 19, pp. 1953–1958, 01 2014.
[18] *Erzeugungsanlagen am Niederspannungsnetz (VDE-AR-N 4105)*, VDE Std., 2018.