UNSUPERVISED ANOMALY DETECTION IN ENERGY GENERATION TIME SERIES USING AN LSTM AUTOENCODER WITH SPATIOTEMPORAL ANALYSIS OF ANOMALIES

Florian Sebastian KIENHOEFER[1*], Johannes SCHOLZ[2]

Introduction, Background and Motivation

The energy generated by photovoltaic (PV) systems varies over time. Distribution system operators (DSOs) record PV output via smart meters and use the resulting time series for billing and analysis. One task is detecting abnormal patterns in the data, often referred to as anomaly detection (AD). For example, these anomalies can be caused by incorrectly wired smart meters or by interpolated values during communication interruptions between DSOs and smart meters. Unidentified anomalies can lead to billing errors and should not be included in energy forecasts, since they would distort them; therefore, it is essential to consistently monitor the recorded data.

A key limitation of existing anomaly detection methods is their lack of analysis of the spatial and temporal patterns of anomalies [1]. This can result in the loss of important information about the origin and spread of anomalies. In addition, not having this information can delay the identification and resolution of grid problems.

To address this gap, we propose a method that uses machine learning to detect abnormal behaviour in PV generation data. We then apply spatial statistics to analyse whether the detected anomalies are spatiotemporally clustered and to identify hotspots that indicate localised issues. We demonstrate our method in a real-world case study.

In this paper we formulate the following research question: How can machine learning and spatial statistical methods be combined to detect anomalies in energy generation time series and assess spatiotemporal clusters?

Methodological Approach

In our approach we rely on a two-stage approach and use an LSTM autoencoder [2] for anomaly detection and Getis–Ord Gi* spatial statistics [3] to assess the spatial distribution at the substation level and to identify anomaly clusters. For the training and evaluation of our approach, we use real-world energy generation time series provided by an Austrian energy distribution system operator.

Autoencoders consist of two parts: Encoders that map the input data to a latent space, which is a compressed representation of it and decoders that map back from the latent space to the input space. In anomaly detection tasks, autoencoders focus on normal patterns in the time series and try to reconstruct them as accurately as possible to the input sequence through minimising a loss function that measures the reconstruction error [4]. Anomalies are usually rare in training data; therefore, the model can reconstruct normal patterns well but struggles with data that contain anomalies. As a result, if the reconstruction error is high, it indicates the presence of an anomaly. Autoencoders are well suited for our task as they are unsupervised and do not require labelled input data for training [5]. Further, they can perform pattern-based anomaly detection, which is ideal for learning the complex, distinct curve trends in energy generation data caused by weather effects or seasonal variations.

For the analysis of spatial anomaly clusters, we chose the Getis–Ord Gi* statistic because it allows us to statistically validate observations that would otherwise rely on visual inspection of spatiotemporal patterns. Getis–Ord Gi* identifies hotspots, where substations with an unusually high number of anomalies are located close together. This suggests a localised, non-systemic issue with a common

[1*] Florian Sebastian Kienhoefer Paris Lodron University of Salzburg, Schillerstraße 30, 5020 Salzburg, floriansebastian.kienhoefer@plus.ac.at, zgis.at
[2] Johannes Scholz Paris Lodron University of Salzburg, Schillerstraße 30, 5020 Salzburg, johannes.scholz@plus.ac.at, zgis.at

cause, rather than a general system-wide problem in the area. It also identifies cold spots, where substations with an unusually low number of anomalies are spatially clustered. This indicates a stable area. Further, it distinguishes these hotspots and cold spots from individual substations with high or low anomaly counts that are isolated coincidences, not reflected in nearby stations and therefore do not form significant clusters.

Conclusions & Implications

Our paper presents an approach to rethink how we preserve and analyse anomalies in time series, emphasising that all data dimensions shall be considered. AI-based methods often neglect the spatial dimension. However, to get a full picture, we advocate utilising all variables and dimensions alike. Including the spatial dimension can offer valuable insights that might be missed if it is disregarded – especially as the spatial domain is rarely evenly distributed, as stated by Tobler's First Law of Geography [6, 7]. Our approach can help develop a more holistic understanding of problems that affect multiple locations. Identified anomaly clusters can indicate systemic issues in the energy grid, equipment failures, or regional disruptions.

References

[1] S. Guan, B. Zhao, Z. Dong, M. Gao, and Z. He, "GTAD: Graph and temporal neural network for multivariate time series anomaly detection," *Entropy,* vol. 24, no. 6, p. 759, May 2022, doi: 10.3390/e24060759.

[2] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff, "LSTM-based encoder-decoder for multi-sensor anomaly detection," arXiv preprint arXiv:1607.00148, Jul. 2016.

[3] A. Getis and J. K. Ord, "The analysis of spatial association by use of distance statistics," *Geographical Analysis,* vol. 24, no. 3, pp. 189–206, Jul. 1992, doi: 10.1111/j.1538-4632.1992.tb00261.x.

[4] J. Pereira and M. Silveira, "Unsupervised anomaly detection in energy time series data using variational recurrent autoencoders with attention," *in 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA),* Dec. 2018, pp. 1275–1282, doi: 10.1109/ICMLA.2018.00207.

[5] O. I. Provotar, Y. M. Linder, and M. M. Veres, "Unsupervised anomaly detection in time series using LSTM-based autoencoders," in *2019 IEEE International Conference on Advanced Trends in Information Theory (ATIT),* Dec. 2019, pp. 513–517, doi: 10.1109/ATIT49449.2019.9030505.

[6] W. R. Tobler, "A computer movie simulating urban growth in the Detroit region," *Economic Geography,* vol. 46, pp. 234–240, Jun. 1970, doi: 10.2307/143141.

[7] H. J. Miller, "Tobler's first law and spatial analysis," *Annals of the Association of American Geographers,* vol. 94, no. 2, pp. 284–289, Jun. 2004, doi: 10.1111/j.1467-8306.2004.09402005.x.