A Commissioning-Oriented Fault Detection Framework for Building Heating Systems Using SARIMAX Models

Parantapa Sawant and Ralph Eismann

Institute for Sustainability in Energy and Construction (INEB),
University of Applied Sciences Northwest Switzerland (FHNW)
parantapa.sawant@fhnw.ch

Abstract. A scalable and rapidly deployable fault detection framework for building heating systems is presented. Unlike existing data-intensive machine learning approaches, a SARIMAX-based concept was implemented to address challenges with limited data availability after commissioning of the plant. The effectiveness of this framework is demonstrated on real-world data from multiple solar thermal systems, indicating potential for extensive field tests and applications for broader systems, including heat pumps and district heating.

Keywords: Building Technologies; Data-Driven Fault Detection; SARIMAX.

1 Introduction

From 2019 to 2023, LoRaWAN-based temperature sensors were installed to monitor the outlet of approximately 450 building solar thermal systems [1]. We developed a rule-based algorithm (RBA) for their fault detection and diagnosis (FDD) by leveraging extensive operational expertise and achieved around 95% accuracy. However, the RBA faced scalability issues due to the variations in plant characteristics, dependent on type of installation or control strategies amongst other factors, that cannot be captured with just one sensor per plant. While it met practical requirements, our current goal is to develop a complementary data-driven algorithm capable of rapid fault detection (FD) across various installations, and potentially extendable to heat pumps and district heating systems. This work presents the results of the first proof-of-concept. The ability of machine learning (ML) algorithms for FD in solar thermal systems is well documented [2], [3], [4]. These studies employed extensive process-history data from multiple sensors, simulation data, and used complex models such as random-forest-regression to detect numerous faults. We also attempted to train ML models using results of the RBA but were hindered by a lack of labeled data and the intrinsic limitations of training ML models to replicate the RBA [5]. Instead, we propose a novel approach utilizing a time series forecasting model that strikes a balance between scalability and minimal data available post-commissioning, facilitating preliminary FD using only the single data-point per installation.

2 Methodology

Conditional data imputation with linear interpolation was performed and only sensors with adequate data quality were chosen for analysis. The *Box-Jenkins Workflow* was then implemented. A daily seasonality and a strong positive linear correlation between the outlet temperature and the global irradiation was identified and hence, a SARIMAX model was implemented. The *statsmodels* package was used in *Python* and hyper-parameters were identified using the ACF-PACF plots as well as the AIC-BIC tests. The number of past days to be used for training was also analyzed empirically to balance accuracy and computational complexity. The results of this manual approach were compared to the automated procedure for hyper-parameter tuning using the *pmdarima* package. The fitted model was used to forecast the target day, which was then labelled as a "fault-day" (F-Day) or "no-fault day" (NF-Day) depending on a threshold for accuracy. The above steps were integrated into a ML-pipeline which was executed once per day, per plant and past NF-Days were used for training in the next iteration.

3 Results

Fig. 1 is an example of the ML-pipeline deployed iteratively to analyze three target days with temperature data from three NF-Days used for training. In the first iteration, data from 19th July to 21st July was used as past data to fit a SARIMAX model which forecast the behavior of the outlet temperature for 22nd July. As the deviation from measured data was within limits it was labelled a NF-Day. Next, for 23rd July, past three NF-Days i.e., 20th July to 22nd July were used. Finally, 25th July was analyzed. It was identified as a F-Day due to unacceptable deviation of the solar outlet temperature from expected no-fault behaviour under similar circumstances. The deployment of this ML-pipeline in field tests to generate alarms based on consecutive FDays is planned after its critical evaluation and streamlining.

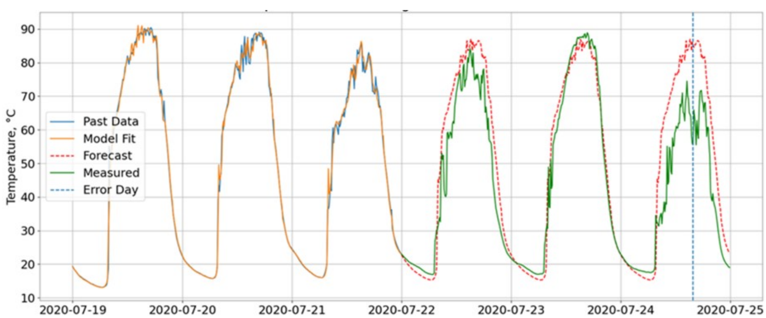

Fig. 1. Results for three iterations of the ML-Pipeline

References

1. Sawant, P., Sintzel, B., Eismann, R., Hofmann, J.W., Sitzmann, B.: Low cost monitoring thermischer solaranlagen mit IoT-sensor und maschinellem lernen [in german only]. available: https://www.aramis.admin.ch/texte/?projectid=49446
2. Faure, G., Vallée, M., Paulus, C., Tran, T.Q.: Fault detection and diagnosis for large solar thermal systems: A review of fault types and applicable methods. Solar Energy **197** 472–484
3. Feierl, L., Unterberger, V., Rossi, C., Gerardts, B., Gaetani, M.: Fault detective: Automatic fault-detection for solar thermal systems based on artificial intelligence. Solar Energy Advances **3** 100033

4. Liu, Z., Liu, Y., Zhang, D., Cai, B., Zheng, C.: Fault diagnosis for a solar assisted heat pump system under incomplete data and expert knowledge. Energy **87** 41–48
5. Hofmann, J.W., Sitzmann, B., Dickinson, J., Kunz, D., Eismann, R.: Use of machine-learning for monitoring solar thermal plants. Journal of Physics: Conference Series **2042**(1) 012007 Publisher: IOP Publishing.