Identifying faults in the building system based on model prediction and residuum analysis

Michael Parzinger¹, Ulrich Wellisch¹, Lucia Hanfstaengl^{1,*}, Ferdinand Sigg¹, Markus Wirnsberger¹, and Uli Spindler¹

¹Rosenheim Technical University of Applied Sciences

Abstract. The energy efficiency of the building HVAC systems can be improved when faults in the running system are known. To this day, there are no cost-efficient, automatic methods that detect faults of the building HVAC systems to a satisfactory degree. This study induces a new method for fault detection that can replace a graphical, user-subjective evaluation of a building data measured on site with an automatic, data-based approach. This method can be a step towards cost-effective monitoring. For this research, the data from a detailed simulation of a residential case study house was used to compare a faultless operation of a building with a faulty operation. We argue that one can detect faults by analysing the properties of residuals of the prediction to the actual data. A machine learning model and an ARX model predict the building operation, and the method employs various statistical tests such as the Sign Test, the Turning Point Test, the Box-Pierce Test and the Bartels-Rank Test. The results show that the amount of data, the type and density of system faults significantly affect the accuracy of the prediction of faults. It became apparent that the challenge is to find a decision rule for the best combination of statistical tests on residuals to predict a fault.

1 Introduction

In the effort to fight global warming, one of the goals of the German government is to achieve a climate-neutral building stock by 2050. The policies focus on two strategies, the use of renewable energies and the increase in energy efficiency. Long-term climate neutrality in the building sector can be achieved by reducing energy consumption and expanding the use of renewable energy [1]. The thermal properties of the building envelope, the efficiency of building technology, and the user behaviour significantly influence energy efficiency of a building. In order to take measures to improve existing buildings, it is important to detect the actual energy efficiency of a building. With the use of on-site measurement data, the actual energy consumption can be detected and flaws in energy efficiency identified.

This study focuses on the energy efficiency of HVAC installations. There are two methods for fault detection, measurement can be performed on-site and faults detected afterwards by analysing the data, or the faults are detected in real-time during operation and reported directly to the building technology manager. Fault detection can be performed by comparing the data collected on-site to a model that delivers expected values. The method is trained with simulation data and applied afterwards to data measured on-site.

Within IEA ECB Annex 71 [2] (International Energy Agency's Energy in Buildings and Communities Program; Annex 71 Building Energy Performance Assessment Based on In-situ Measurements) the members of the Annex were to solve an exercise to explore system identification tech-

niques. Compared to the exercise, this study focuses on the faults of the building system. Simulation data of the twin houses, which are two identical case study houses at the Fraunhofer Institute for Building Physics in Holzkirchen, Germany were used. The simulation was carried out within a validation study of the IEA ECB Annex 58 and 71 project [3], [2]. The data consist of two sets, a first part in which fault free operations are simulated and a second part in which various system faults were integrated into the simulation. Both data sets have the length of one month.

In a first phase two different statistical models, a random forest (a machine learning method) and an ARX model (a time series method), predict the normal operation by predicting the total heating power of the building. In the second phase, these two prediction models predict the data set that contains faults. The fault detection is carried out using residual analysis, model checking based on residual analysis is a standard technique for time series analysis, cf. [4], page 175 ff. and [5], page 360 ff. With a suitable time series model adaptation, the residuals are generally assumed to be an approximate white noise and i.i.d. with a mean value of zero. In this study we use this characteristic as the starting point for a decision technique. We propose that the decision method is appropriate for general residuals based on a good model fitting (e.g. resulting from a random forest model).

Fassois and Sakellariou [6] give an overview of time series methods for fault detection in vibrating structures. There are two main types of time series methods for fault detection: non-parametric methods, which use spectral analysis, and parametric methods, which can be categorised as parameter-based and residual-based methods. For the residual technique, the estimations of the model parameters do not have to be considered. The residuals

* Corresponding author: Lucia.Hanfstaengl@th-rosenheim.de

can be calculated directly from the predictions (based on the same modelling method) and the responses. The white noise property of the residuals can be analysed using various statistical tests, which work partially as portmanteau tests. In this study a data-driven decision rule for fault detection merges multiple tests. The approach of combining tests involves a general applicability of the decision rule. Different faults and different prediction methods for the response yield different deviations of the standard behaviour of the resulting residuals. The decision rule for fault is learned on the residual data and can be seen as a sort of portmanteau decision rule for fault detection where the null hypothesis is specified by faultlessness. The technique is adjusted to the situations of observed and unobserved faults in the learning sample. Furthermore, the method is formed for fault detection in specific time points and time intervals.

The developed methods for fault detection could replace a graphical, user-subjective valuation of a residual plot using an automatic, data-based approach.

2 Description of simulated data and system faults

The simulation was build upon an empirical validation experiment of Annex 58 and Annex 71. A detailed description of the two identical full-size buildings of the Fraunhofer Institute for Building Physics in Holzkirchen, Germany, and data of the Annex 58 experiment can be found in [7], [8], [9] and [10].

The data set is obtained through detailed simulation with the program IDA ICE [11]. The simulation uses the house description of Annex 58 and 71 and the climate boundary conditions of January and February 2019 in Holzkirchen (Annex 71).

For the simulation model, each room was equipped with a 2000 W electric radiator with a longwave radiation fraction of 40%. The heating set point of the air temperature was set to 21°C and controlled room wise by a thermostatic control with a dead band of ± 0.5 K. The simulation integrates a MHVR (Mechanical Ventilation with Heat Recovery) air handling unit with a heat recovery of 80 percent with an integrated MVHR summer bypass (possibility to switch off the heat recovery during summer months). The simulation includes a simple occupancy plan of a four-person household with the absence of users between 7:30 and 17:00 each day. The data set starts on January 1st and ends February 28th. The first month (1/Jan - 31/Jan), the building runs in regular operation. The second month (1/Feb - 28/Feb) includes faults in the operation of the building. In this study three faults are selected. The first fault is a circuit breaker failure (F1; Circuit breaker failure of the electrical heating in the upper floor; 2. Feb. 0:00 - 4. Feb. 23:59), the second fault a failure of the MHVRs heat recovery unit (F2; MHVR summer bypass switch off the heat recovery during fault duration; 10. Feb. 0:00 - 15. Feb. 23:59), and the third fault a higher thermostat set point than necessary (F3 - Living room thermostat to set temperature 28°C; 20. Feb. 0:00 - 23. Feb. 23:59). The data set contains the following indoor and outdoor properties. Air temperature for each room and total heating power supplied by all electrical

radiators are measured indoors. The outdoor properties are the air temperature, relative humidity, diffuse and direct solar irradiation on horizontal surfaces, and wind speed and wind direction.

3 Statistical tests

The presented statistical tests are implemented with the programming language R [12], and the user interface RStudio [13]. The graphics were created with the R package "ggplot2" [14].

The predictive models for the response total heating power use as predictors the indoor temperatures of all rooms, all outdoor information and the daytime in hours. The time-dynamic models take the differences between the value of response at the current time and the value at the time one hour ago (time lag one hour) into account. The total heating power is the model output and the values of the total heating power one hour and two hours ago at each point in time are added as predictors (features) to the models. The predictive modelling is carried out with the method random forest [15] and an ARX (autoregressive with exogenous variables) time series model [16]. To predict the total heating power $\hat{y}_t$ in February, the January data is used for training and the February data for testing. The total heating power in January is predicted with 4-fold cross-validation. The 4-fold cross-validation divides the January data into four nearly equally sized parts. Then three of these parts predict the reminding part. The difference between the real values (responses) and the predicted values of the total heating power are the residuals. Let y_t be the response of the observed total heating power and $\hat{y}_t$ the predicted total heating power from a model at the time $t = 1, \dots, n$. Then $\hat{\epsilon}_t := y_t - \hat{y}_t$ denotes the residual at the time t and $\hat{\epsilon} := (\hat{\epsilon}_1, \dots, \hat{\epsilon}_n)$ the vector of the residuals. Furthermore $\tilde{\epsilon} := (\tilde{\epsilon}_1, \dots, \tilde{\epsilon}_n, \tilde{\epsilon}_{n+1}, \dots, \tilde{\epsilon}_{n+L-1}) := (\hat{\epsilon}_1, \dots, \hat{\epsilon}_n, \hat{\epsilon}_1, \dots, \hat{\epsilon}_{L-1})$ is defined for a fixed $L \in \{2, \dots, n\}$.

Figure 1 shows the January and the February residuals for the developed random forest and the ARX model.

After successful modelling, the typical properties of residuals are to be statistically tested. A fault in the data process is assumed if the behaviour of the residuals deviates significantly from the standard properties. The special properties of residuals depend on the modelling methodology, the data structure of the learning sample, as well as on the prediction quality of the model. It is assumed that the residuals have a median of zero, are independent and therefore uncorrelated from each other, and behave randomly. The Sign Test [17] and the Wilcoxon Signed-Rank Test [18] are suitable for testing for median equal to zero. The Turning Point Test [19] is well suited for testing independence, the Box-Pierce Test and the Ljung-Box Test [20] for autocorrelation. Randomness can be tested with the Bartels-Rank Test, Cox-Stuart Trend Test, Difference-Sign Test and Mann-Kendall Rank Test [21]. In total, this study examines the residuals using nine tests divided into four test objectives.

Fig. 1. January and February residuals for ARX model and random forest model.

3.1 Moving p-value

The moving residuals for the shift s with time window length $L \in \{2, \dots, n\}$, which represents the sample size of the moving residuals, are defined by $M_{\tilde{\epsilon}}(s, L) := (\tilde{\epsilon}_{1+s}, \tilde{\epsilon}_{2+s}, \dots, \tilde{\epsilon}_{L+s})$. The moving residuals are used in order to avoid testing all residuals at once. If $p_T(\cdot)$ is the p-value of the statistical test T , it is used $(p_T(M_{\tilde{\epsilon}}(0, L)), p_T(M_{\tilde{\epsilon}}(1, L)), \dots, p_T(M_{\tilde{\epsilon}}(n-1, L)))$, for a fixed L , to examine periods for faults. If the p-value of a test is less than a previously selected significance level $\alpha \in (0, 1)$, then the null hypothesis of this test is significantly not met [5], page 5 ff. For all nine tests the null hypothesis is that a certain property of the residuals is fulfilled. Therefore it applies that for each test a fault is suspected when the p-value of this test is smaller than α .

3.2 Mean p-value (MPV)

A disadvantage of the p-values of the moving residuals used so far is that it can be recognised at which shift s the p-value is no longer as expected, but not at which time point. In the following, a new constructed function determines faulty time points. This is made possible by a mean of the p-values from the moving residuals. Let

$$\Gamma_{T,L}(\hat{\epsilon}_t) = \frac{1}{L} \sum_{s=0}^{n-1} p_T(M_{\tilde{\epsilon}}(s, L)) I_{\{\tilde{\epsilon}_{1+s}, \dots, \tilde{\epsilon}_{L+s}\}}(\hat{\epsilon}_t)$$

be the mean p-value, in future called MPV. Where $I_A(a)$ denotes the indicator function with $I_A(a) = 1$, if $a \in A$ and 0 else. The MPV for a given test T and a window length L is for each $\hat{\epsilon}_t$ the mean of all p-values generated by moving residuals that used $\hat{\epsilon}_t$ in the calculation. For the fault detection it applies, if $\Gamma_{T,L}(\hat{\epsilon}_t) < \alpha$, then the test T assumes a fault at time point t . Fault detection uses a combination of tests, which raises the question of how many tests must assume at least one fault to assume a fault in total. This value is abbreviated with H . The value of H is at least one and at most the whole number of tests which are used in the analysis. In this work nine tests are used. Thus an automated fault detection can be defined by using the three parameters $L \in \{2, \dots, n\}$, $\alpha \in (0, 1)$ and $H \in \{1, 2, \dots, 9\}$. Therefore, in the following each triple $(L, \alpha, H) \in G := \{2, \dots, n\} \times (0, 1) \times \{1, 2, \dots, 9\}$ is simplified termed as a decision rule. For the application G is replaced by a large subset of G . Let

$$\Phi : G \rightarrow \{0, 1, \dots, n\},$$

$$\Phi(L, \alpha, H) \mapsto \text{Number of estimated faults in January}$$

be the fault function. Since there are no system faults implemented in the January data a limitation is made to the decision rules which not erroneously detect faults in the January data. Accordingly, a decision rule from the following set is required which is referred to as choice set.

$$C := \{(L, \alpha, H) \in G \mid \Phi(L, \alpha, H) = 0\},$$

Fig. 2. Fault estimate with decision rule (106, 11.2%, 8) for the ARX model and (397, 6.2%, 7) for the random forest model. The periods with faults in the data are marked in red. The y-axis indicates if the decision rule decided on faults.

Fig. 3. Fault estimation with decision rule (119, 11.8%, 7) for the ARX model and (101, 20%, 8) for the random forest model. The periods with faults in the data are marked in red. The y-axis indicates if the decision rule decided on faults.

Fig. 4. Fault estimation with decision rule (84, 10.4%, 8) for the ARX model and (97, 9.6%, 7) for the random forest model. The periods with faults in the data are marked in red. The y-axis indicates whether the decision rule decided on faults.

Fig. 5. Fault estimate with decision rule (392, 3%, 7) for the ARX model and (545, 10.6%, 7) for the random forest model. The periods with faults in the data are marked in red. The y-axis indicates whether the decision rule decided on faults.

3.3 Parameter optimisations in the case of observed faults

A grid search finds with the known time points of faults an optimised decision rule $(L, \alpha, H) \in C$, which minimises a previously defined fault rate. This optimisation shows on the data of this study good results. The applicability of the model, which was optimised by this procedure on another data, is not tested so far. A problem could be that the decision rule adapts to the data too individually. In future works, the procedure should be repeated with other validation data. Only when data with known faults is available, this technique can be used, which is why future works investigate other techniques for data with unknown faults or without faults.

4 Results

The January residuals, created by a 4 fold cross-validation, are useful to compare both prediction models. This can be done by using the mean squared error (MSE) of each model, which is defined by $\frac{1}{n} \sum_{t=1}^n \hat{\epsilon}_t^2$. For the ARX model, the calculated MSE equals 31,001.55, and for the random forest, it is 100,274.7. For this reason, the ARX model seems to be a better choice to predict the total heating power. We question if a better prediction model is automatically better for fault detection?

Concerning the case that the analysis observed the faults, and the times at which the faults occur are known, the decision rule can be optimised accordingly. This research uses the decision rules of the choice set for each data set and each model with the lowest prediction error rate.

This study analysed whether the ARX model or the random forest model delivers better results when the first, the second or the third fault or all three faults are in the data. Figures 2,3 and 4 show the results when only adding one fault to the data. If only the third fault is in the data the decision rule using the ARX model delivers better result because the decision rule based on the random forest model often suspects faults where no faults are and also matches the existing fault less accurately. If all faults are in the data as shown in Figure 5, both models detect the third fault very well, the ARX model has problems with the second fault, and both models recognise the first fault not well.

5 Conclusion

This study uses residual analysis for fault detection of HVAC systems in buildings. A detailed simulation of a residential case study house provided the data for the analysis. The predictive modelling of the total heating power was carried out with the method random forest and ARX (autoregressive with exogenous variables) time series model. The residuals were calculated directly from the predictions. A combination of statistical tests explored the white noise properties of the residuals, while a data-driven decision rule that combines multiple tests predicted the faults. The methods for fault detection developed in this study could replace a graphical, user-subjective evaluation of a residual plot using an automatic, data-based approach.

A fault detection method that uses residual analysis has several advantages: For one the method does not depend on the kind of prediction model applied, furthermore information such as model parameter estimations and specific model structures are become superfluous. The research has proved that the methods of fault detection can be applied to a wide variety of data and prediction models such as time series models and procedures of machine learning (including black-box methods). Statistical tests can be added or removed depending on the suspected residual properties.

This study introduces a decision rule when faults are observed. The method depended heavily on the p-values, which usually depend on the sample size, and the methods had to be consistently adjusted to the specific situation and the given sample size. A method that investigates data when faults are unobserved is part of future research. Future studies can avoid the dependency on the p-values by using the test statistics it selves instead of the p-values and defining corresponding threshold values, or by using tests and methods of the Bayesian statistics. Another possible improvement could be to integrate statistical tests that apply the frequency characteristics of the residuals time series.

The method invites the application on data measured on-site. A real building evaluation could be performed using a learning data set based on a building simulation (with simulated, observed faults) of the same building. Consequently, it is essential to test the applicability of a decision rule based on simulated building data on the behaviour of the original building. The accuracy of the building simulation would then be a significant factor in the successful application of the method. Another practical application is the development of a decision rule for fault detection based on real building data set (with enforced, observed faults).

This study used a simulated data set of two months. A future study with a simulated data set with data over one year is planned. In the case that no simulated data are available, but a data set of in situ data, a one month "error-free" data set could be sufficient to determine a significant prediction model. In a further study of Annex 71 [2] in spring 2020, this is the task and the methods developed in this study are to be tested.

Supported by: Federal Ministry for Economic Affairs and Energy on the basis of a decision by the German Bundestag. projectnumber: FKZ03ET1509C

References

- [1] BMWi, Energieeffizienzstrategie Gebäude. Wege zu einem nahezu klimaneutralen Gebäudebestand (2015), <https://www.bmwi.de/Redaktion/DE/Publikationen/Energie/energieeffizienzstrategie-gebäude.html>
- [2] IEA ECB Annex 71. Building Energy Performance Assessment Based on In-situ Measurements, Website (2016-2021), <https://www.ecbcs.org/projects/project?AnnexID=71>

- [3] IEA ECB Annex 58. Reliable Building Energy Performance Characterisation Based on Full Scale Dynamic Measurements (2011-2016), <https://www.iea-ebc.org/projects/project?AnnexID=58>
- [4] H. Madsen, *Time series analysis*, Chapman and Hall (2008)
- [5] H. Pruscha, *Statistisches Methodenbuch* (Springer, 2006)
- [6] S.D. Fassois, J.S. Sakellariou, *Time-series methods for fault detection and identification in vibrating structures*, Philosophical Transactions of the Royal Society, Vol. **77**, pp. 411–448 (2006)
- [7] P. Strachan, K. Svehla, I. Heusler, M. Kersken, *Whole model empirical validation on a full-scale building*, Journal of Building Performance Simulation, Vol. **9**, pp. 331–350 (2016)
- [8] M. Kersken, I. Heusler, P. Strachan, *Erstellung eines neuen, messdatengestützten validierungsszenarios für gebäudesimulationsprogramme*, Fifth German-Austrian IBPSA Conference at RWTH Aachen University, pp. 144–151 (2014)
- [9] P. Strachan, *Twin houses empirical dataset: Experiment 1* (2015), <https://pureportal.strath.ac.uk/en/datasets/twin-houses-empirical-dataset-experiment-1>
- [10] P. Strachan, *Twin houses empirical validation dataset: Experiment 2* (2016), <https://pureportal.strath.ac.uk/en/datasets/twin-houses-empirical-validation-dataset-experiment-2>
- [11] IDA Indoor Climate and Energy (IDA ICE), simulation tool (2019), <https://www.equa.se/en/ida-ice>
- [12] R Core Team, R version 3.6.1 (Action of the Toes), <https://www.r-project.org/>
- [13] RStudio Team, RStudio: Integrated Development Environment for R (2019), <http://www.rstudio.com/>
- [14] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, 2016)
- [15] M.N. Wright, A. Ziegler, *ranger: A fast implementation of random forests for high dimensional data in C++ and R*, Journal of Statistical Software, Vol. **77**, pp. 1–17 (2017)
- [16] K. Ohtsu, H. Peng, G. Kitagawa, *Time series modeling for analysis and control*, Springer (2015)
- [17] A.T. Arnholt, B. Evans, *Bsda: Basic statistics and data analysis* (2017), r package version 1.2.0 <http://CRAN.R-project.org/package=BSDA>
- [18] R Core Team, *Wilcoxon rank sum and signed rank tests*, <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/wilcox.test.html>
- [19] A. Hart, S. Martínez, *spgs: Statistical patterns in genomic sequences* (2018), r package version 1.0-2 <http://CRAN.R-project.org/package=spgs>
- [20] R Core Team, *Box pierce and ljung box tests*, <https://stat.ethz.ch/R-manual/R-patched/library/stats/html/box.test.html>
- [21] F. Caeiro, A. Mateus, *randtests: Testing randomness in r* (2014), r package version 1.0 <http://CRAN.R-project.org/package=randtests>