An easy to set up residual generator based on multilayer perceptron networks and Bayesian optimisation for the application in automated fault detection and diagnosis in building systems

Sebastian Dietz[1], Frank Scholzen[1], Nicolas Réhault[2], Cédric Dockendorf[3]
[1]University of Luxembourg, Luxembourg, Luxembourg
[2]Fraunhofer-Institut for Solar Energy Systems ISE, Freiburg, Germany
[3]Symvio S.à.r.l, Strassen, Luxembourg

Abstract

Automatic fault detection and diagnosis (FDD) methods are rarely used in building systems due to their individual design. We present a residual generating FDD approach combining multilayer perceptron networks trained with historical data and Bayesian optimisation for hyperparameter tuning. A comprehensive engineering process has been developed, which is highly automated and applicable by non-machine learning experts. We demonstrate the transferability using datasets from twelve different air handling units and provide an estimation of fault-free behaviour. Applied on a synthetic data set, the approach shows comparably results to a rule-based fault detection, with the advantages of less threshold tuning, detecting unknown faults, and facilitating fault diagnosis based on residuals.

Highlights

- Machine learning based fault detection for HVAC
- Residual generating approach enables fault diagnosis
- Applicable by non-machine learning experts
- Highly automated and transferable
- Demonstrated on twelve data sets from different AHUs

Introduction

Buildings are responsible for about 36% of the global final energy consumption and 39% of the CO_2 emissions (IEA, 2019). At the same time, inefficient or faulty operation of Heating, Ventilation and Air Conditioning (HVAC) systems is a common issue (Réhault, 2010). Automated fault detection and diagnosis (FDD) methods can effectively support continuous commissioning of buildings and achieve benefits such as energy savings, improved room comfort, and better maintenance staff scheduling. However, despite intensive research, FDD methods are rarely implemented for HVAC systems in buildings (Matetić et al., 2023).

Numerous FDD approaches for HVAC systems are presented in the literature (Melgaard et al., 2022). Matetić et al. (2023) categories them in knowledge discovery (e.g., rule based), physics based, and data-driven approaches. However, regardless of category, accurate modelling is crucial for FDD. To achieve this, it is very challenging and time-consuming to model a dynamic non-linear system. Rule-based methods require extensive fine-tuning to determine suitable thresholds, while in the case of physical models a deep understanding of the process, a complex modelling, and an individual calibration of the model parameters is needed. These challenges pose significant obstacles to apply existing FDD methods in buildings, where HVAC systems are individually designed to the respective building energy concept and user requirements. The development and implementation costs are often too high to be refinanced through the savings. In contrast, data-driven approaches can substantially reduce modelling efforts by learning system behaviour from historical operational data (Mirnaghi et al., 2020).

The FDD process is composed of two subtasks: fault detection and fault diagnosis. The latter includes fault isolation and fault identification (Katipamula, 2005). Fault detection (FD) can be described as a two-class classification problem (fault-free / faulty) (Benndorf et al., 2018). For this task, unsupervised or supervised machine learning (ML) techniques are frequently used to detect anomalies in operating data. Although this approach is highly flexible and effective, as it can detect unknown faults, it has the drawback of lacking information about the fault source and size. Thus, it cannot support fault diagnosis, which is an important feature in facility management applications.

A general problem to implement ML methods in the FDD process for HVAC systems is the lack of faulty operational data. Thus, it is not possible to extensively learn all possible faulty states directly from historical data. In residual generating approaches, in contrast, the normal behaviour is estimated, and the detection and diagnostic of faults is based on the deviations of the observed system variables. For this task, state space models (so-called observers) are usually used in the field of control engineering. However, a promising approach is to replace the observer with suitable ML models.

Related work to residual generating FDD approaches involving ML methods can be found in Wang et al. (2016). They present a robust fault detection and diagnosis strategy for multiple faults in ventilation systems. While the fault detection is based on a statistical model and the resulting residuals, the diagnosis is carried out using expert rules. A similar approach is taken by Liao et al. (2021) by combining a Convolutional Neural Network with a rule-based system. A feed forward neural network (NN) combined with rules for fault diagnosis on subsystem level of an air handling unit (AHU) is used by Lee et al. (2004). Bezyan et al. (2022) apply support

vector machine models to predict temperatures of two target sensors in an AHU, while a second recurrent neural network is used to predict the regressor inputs under normal conditions. Fault diagnosis is done by rule-based techniques. Du et al. (2014) focus on faults in heating coils and combine a basic NN and an auxiliary NN to detect abnormities and classify the fault by a subtractive clustering analysis. A distributed concept at the component level of a ventilation system (cooling coil and VAV box) is pursued by Shahnazari et al. (2019). For this, they use recurrent neural networks to predict the fault-free behaviour. The presented approaches allow diagnoses of multiple faults via expert rules.

However, the transferability has not been sufficiently investigated in the presented methods. It is known that the performance of ML methods heavily relies on the given training data set. Additionally, the process must adapt automatically to the new system's conditions (e.g., number of data points).

In this paper, a comprehensive engineering process to set up a residual generating FD process has been developed, with the goal to reduce the implementation effort to a minimum by applying well-established methods from the field of ML. The process is highly automated and applicable by non-machine learning experts (e.g., HVAC engineers). Nominal behaviour of the observed system is estimated by a bank of multilayer perceptron networks and Bayesian optimisation for hyperparameter tuning. Based on the results, residuals are calculated for each observed variable and fault detection with a threshold method is applied. Furthermore, the resulting residual patterns facilitate fault diagnosis, which will be in the focus of future works. To investigate the transferability, the process is applied to a total of twelve AHUs from two different buildings and the quality of estimate for residual generation is studied, while fault detection is applied on a simulated data set with known fault states. Although in this work only AHUs are considered, the method can be applied to other parts of HVAC systems in buildings.

Description of the process

The basis for a data driven FDD process are time-series data from the building automation system (BAS). In this work, the control loops of the AHU's serve as the system boundaries for selecting the relevant data points. In one of the case studies, this also includes the pump in the water circuit located adjacent to the AHU. To allow an automatic pre-processing of raw time-series data the user must provide basic metainformation (Figure 1). This metadata includes information for each datapoint about unit, range of values, scaling factor, offset and the sample method (average / difference). But most important is a mapping of the origin names from the BAS to standardised semantics scheme (e.g., “supa_mea_t” for the measured time series of the supply air temperature). In addition, the user must label fault-free time periods for training and validation. This can be checked manually or with the help of well-known rules. However, the true state is difficult to prove. Ideally, the training and validation period should cover one year.

Data pre-processing & feature selection

Pre-processing is an important part of training ML models as it can significantly influence the quality of the results. Because data from BAS can be heterogeneous depending on the source, conventional data handling functions must be performed. This step involves: removing duplicates, sorting data over time, cleaning counter data, scaling, and offsetting, converting units, limit checking, and sampling data into a uniform time interval. Interpolation of data should be avoided as far as possible, as it can result in the model learning the wrong behaviour. However, regular data gaps in BMS records are common, and one single sensor affected by frequent gaps can already significantly reduce the amount of data available for model training, in which case interpolation should be considered.

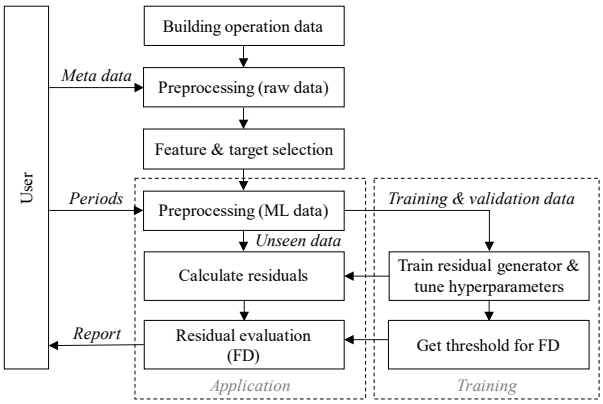

Figure 1: Scheme of the implemented FD-process.

Feature selection methods reduce the number of input variables of a ML model and thus its complexity. In this work, an extendable set of system-specific expert rules is applied to filter individual variables. For instance, the control signal for the filling valve of the adiabatic cooling system contains no information to estimate the states in the air stream and can only be estimated itself if the fill level of the tank is recorded. Additionally, a steady-state detector removes constant data points such as set points with no changes over time within the training period.

Time series data can be categorical or continuous. Even though most operational data in buildings are continuous, some, like signals from pumps or dampers, have categorical properties. When categorical data gets detected, a classification model instead of a regression model is trained to estimate the target, as it is typically more effective in this case. Regardless of the model type, the input data is transformed into a sliding time window to incorporate system dynamics. In this study, a time window of two hours is employed. If $\vec{x}(t)$ is a vector with operating data at time *t* with *m* elements, then the input vector $\vec{x}_w(t)$ is constructed from a window with *n* records according to Formula (1).

$$\vec{x}_w(t) = [x_1(t), x_1(t-1), \dots, x_1(t-n+1), \\ x_2(t), x_2(t-1), \dots, x_2(t-n+1), \dots, \\ x_m(t),\ x_m(t-1), \dots, x_m(t-n+1)] \qquad (1)$$

When the system is off, the characteristic of the observed data changes significant. During these periods, the state variables are influenced more by unknown boundary

conditions rather than the system itself. To overcome this issue, only data from periods where the system is switched on is used. Finally, to prepare the input data for the model training, a min-max scaler is employed to normalise the data into a range between 0 and 1.

Estimation of nominal behaviour

To estimate the actual states and control signals for a normal operation of the system, a bank of regression and classification models is constructed. For the estimation of the i-th state variable v_i, all state variables $\vec{v}$ except the i-th element, here denoted as $\tilde{v}$, and all control signals $\vec{u}$ are given as input variables. The control signal u_i is estimated accordingly, where $\tilde{u}$ denotes the control signal vector without the i-th element (Formula (2)).

$$v_i = f(\tilde{v}, \vec{u}); u_i = f(\tilde{u}, \vec{v}) \quad (2)$$

For each estimator, a multilayer perceptron neural network is trained. While for regression targets a NN with ReLU activation functions and drop out in all layers is constructed, for categorical data a classifying MLP with one hot encoding and Softmax activation function for the output layer is trained. Table 1 gives an overview of the hyperparameters used in this work. To prevent model overfitting on training data, the learning process stops before the maximum number of epochs is reached, when the validation loss makes no progress.

Table 1: Hyperparameters of the MLP NN. Range based values are selected by applying Bayesian optimisation.

Hyperparameter	Value
Maximum number of epochs	60
Batch-size	30
Learning rate	$1e^{-4}$, $1e^{-5}$
Loss function regression	RMSE
Loss function classification	Cross-entropy
Number of hidden layers	3-6, step 1
Number of neurons in each layer	24-552, step 48
Dropout first hidden layer	0,1
Dropout 2nd to n hidden layer	0,2
Activation function	ReLu
Scaling method	Min-max scaling
Encoder (classification)	One-hot encoder

It should be noted that any type of ML model can be used if the nominal behaviour of the system is estimated with sufficient accuracy. However, finding a suitable model structure and hyperparameters is a challenging and time-consuming task in praxis. To overcome this problem, Bayesian optimisation is used to find an appropriate set of parameters. For regression, the objective is to minimise the root mean squared error (RMSE) within the validation period. In Formula (3), $y(t)$ denotes the observation, $\hat{y}(t)$ the estimated value, and n the number of values in the validation period.

$$\arg\min\left(\sqrt{1/n \sum_{t=1}^{n}(y(t) - \hat{y}(t))^2}\right) \quad (3)$$

In classification problems, RMSE is not applicable, and instead the objective is to reduce cross-entropy losses. Formula (4) shows the objective function where $q_j(t)$ is the true one hot encoded value of the t-th observation and the j-the class. The estimated probability for the j-th operation and t-th class denotes $p_j(t)$. The number of observations in the validation period is n and m the number categorical classes.

$$\arg\min\left(-1/n \sum_{t=1}^{n}[\sum_{j=1}^{m} t_j(\mathrm{t}) < \log(p_j(\mathrm{t}))]\right) \quad (4)$$

Residual calculation & evaluation

The bank of MLP models estimates all state and control variables $\vec{\hat{y}}(t)$ at timestep t. Together with the observations $\vec{y}(t)$ the residuals $\vec{r}(t)$ are calculated according to Formula (5).

$$\vec{r}(t) = \mathrm{MinMax}(-1,1,(\vec{y}(t) - \vec{\hat{y}}(t))) \quad (5)$$

It is important to mention, that the calculation is performed using scaled values between 0 and 1 to ensure that the deviation's magnitude is not affected by the original variable range. However, since we are dealing with black box models, it is possible that the limits will be exceeded in the event of a fault or in unknown operating conditions. The exceedance has no physical significance. So, for FD the residuals are restricted to a range between -1 and 1.

The L^2 norm of the residual vector in Formula (6) has proven to be suitable to be evaluated for FD (Ding, 2013). The required threshold value is obtained from the standard deviation determined in fault-free operation, such as the training or validation period. If J_{th} denote a threshold and f a scale factor, then the detection logic in Formulas (7) and (8) results.

$$\|\vec{r}(t)\|_2 = \sqrt{\sum_{i=1}^{n}(r_i(t))^2} \quad (6)$$

$$J_{th} = \mathrm{f} * \mathrm{std}_{fault\ free}(\|\vec{r}\|_2) \quad (7)$$

$$FD = \begin{cases} False; & \|\vec{r}(t)\|_2 \le J_{th} \\ True; & \|\vec{r}(t)\|_2 > J_{th} \end{cases} \quad (8)$$

Due to the standard deviation as a threshold value, model uncertainty is automatically accounted in FD. However, this implies also that the worse the model is, the more the system becomes tolerant to faults. The sensitivity of FD can be adjusted by the scale factor, whereby a factor of $f = 2$ is applied in this work.

Description of the case studies

The residual generator's performance is evaluated using data from twelve AHUs from two different buildings. The system utilised the operational data provided by the BAS without any additional sensors being installed. Nevertheless, due to the absence of labelled data, a synthetic data set was generated to assess FD.

Case study MSA

The Maison du Savoir (MSA) building houses the University of Luxembourg. A total of six ventilation systems are observed. Each system is designed to supply fresh air, heating, and cooling to a specific auditorium. Structurally all six systems are identical and equipped with a heating coil, adiabatic cooling, a heat recovery wheel, filters, and dampers which allow operation in return air mode. In cooling mode, water droplets are introduced into the extracted air stream and subsequently evaporated. This evaporation causes a reduction in temperature, enabling the transfer of cooling energy to the

supply air stream through heat recovery. A district heating system supplies the system with heat. The systems vary in terms of nominal volume flow rates and subcomponent sizing depending on the size of the auditorium. The air volume flow is controlled in two stages (Table 2) with respect to the actual needs of the zone. Ongoing data acquisition started in May 2021 with an interval of five minutes. Due to a significant modification in the control strategy that involved variable air flow based on CO_2 measurements, only data up to mid-October 2022 is considered and divided as follows:

- Training period: 2021-05-19 to 2022-01-18
- Validation period: 2022-01-19 to 2022-07-18
- Test period: 2022-07-19 to 2022-10-18

The data points used to build the residual generator shows Table 4. For clarity, data points that are recorded but discarded during feature selection are not shown.

Table 2: Nominal air volume flows of the observed AHUs in the MSA building.

Name(s) of AHU(s)	Volume flow (60%)	Volume flow (100%)
AUD-2 & AUD-11	3150 m^3/h	5250 m^3/h
AUD-4 & AUD-8	1890 m^3/h	3150 m^3/h
AUD-7 & AUD-9	5040 m^3/h	8400 m^3/h

To produce a synthetic data set of labelled data with known faulty states, a physical model of AHU AUD8 is created within the TRNSYS simulation environment. This data set replaces real building operating data, to evaluate the entire FD process shown in Figure 1.

The physical model is adjusted as far as possible to the real system. To account for potential variations in climate conditions, the model is simulated using two different weather data sets from the same climate zone (Germany). Additionally, three different fault types are implemented with random fault size and occurrence times. The first fault pertains to a stuck control valve within the heating coil circuit (actor fault), the second fault relates to an offset in the temperature sensor in the extract air stream (sensor fault), and the third fault considers downtimes for the heat recovery wheel (process fault). A total of four years of simulations are conducted, encompassing both fault-free and faulty behaviour, which are then employed to train and assess residual generation as follows:

- Training: One year, fault free, weather data 1
- Validation: 1st half of a year, fault free, weather data 2
- Testing: 2nd half a year, fault free, weather data 2
- Evaluation of fault detection: Two years with random faulty states, weather data 1 and 2

Case study PSHN

The second case study is a primary school in Hohen Neuendorf (PSHN), Germany. A total of six AHUs are observed (Table 3), whereby each AHU serves a zone composed of six to eight classrooms, corridors, and sanitary rooms. Exceptions are the AHUs for the assembly hall (Asse-Hall) and the sports hall (Spo-Hall). A hybrid ventilation system has been implemented, incorporating both mechanical ventilation and natural ventilation through motorised openings (Dietz, 2016). The AHUs operate in two stages, with the second stage being activated in extreme hot or cold weather conditions when natural ventilation openings don't open. A heating coil, a plate heat exchanger for heat recovery, outside air dampers and filters are installed. Additional radiators within the zone control room air temperature. A central pellet boiler serves the building with heat. Data in a five-minute interval of the following periods are used:

- Training period: 2015-01-01 to 2015-12-31
- Validation period: 2016-01-01 to 2016-12-31
- Test period: 2017-01-01 to 2017-12-31

Table 3: Nominal air volume flows of the observed AHUs in the PSHN building.

Name(s) of AHU(s)	Volume flow stage 1	Volume flow stage 2
Zone-A	1650 m^3/h	3150 m^3/h
Zone-B & Zone-C	950 m^3/h	2850 m^3/h
Zone-D	1430 m^3/h	2640 m^3/h
Asse-Hall	1500 m^3/h	4500 m^3/h
Spo-Hall	1890 m^3/h	3570 m^3/h

Table 4: Description of the input data for the generation of residuals. "x" indicates whether the data point is available for the case study.

Symbol	Description	MSA	PSHN
T_{supa}	Supply air temperature	x	x
rH_{supa}	Supply air humidity	x	(x)[1]
T_{exha}	Exhaust air temperature	x	x
rH_{exha}	Exhaust air humidity	x	(x)[1]
$T_{supa,ahrc}$	Supply air temperature after heat recovery	-	-
$T_{exha,aadc}$	Exhaust air temperature after adiabatic cooling	x	-
T_{room}	Room air temperature	x	-
T_{oa}	Outside air temperature (in duct)	x	x
rH_{oa}	Outside air humidity (in duct)	x	x
T_{exhao}	Exhaust air out temperature	x	x
rH_{exhao}	Exhaust air out humidity	x	(x)[1]
U_{recirc}	Control signal of return air damper	x	-
$U_{supa\text{-}fan}$	Control signal of supply air fan	x	-
V_{supa}	Supply air volume flow	-	x
$U_{exha\text{-}fan}$	Control signal of exhaust air fan	x	-
V_{exha}	Exhaust air volume flow	-	x
U_{hrc}	Control signal of the heat recovery wheel	x	-
$U_{hrc,byp}$	Damper signal for bypass of heat recovery	-	x
$U_{adc,pu}$	Signal for pump in the adiabat cooling system	x	x
$U_{hc,pu}$	Signal for pump in the heating coil circuit	x	x
$T_{hc,supw}$	Supply water temperature in the heating coil circuit	-	x
$T_{hc,retw}$	Return water temperature in the heating coil circuit	x	x
$U_{hc,valve}$	Control signal for the valve in the heating coil circuit	x	x
[1] *only for AHU Asse-Hall*			

Results

Considering the objective of creating an FD process that is highly automated, the presented residual generator is applied on each data set of the case studies without modifying the parameters individually. The estimation accuracy is evaluated using a test data set that is not seen during the training process. Additionally, the performance of FD is assessed based on the synthetic data set with known fault states.

Quality of estimate

It is important to recall, that for model training, a fault-free operation data set is assumed. However, for operating data of a real system the true states are unknown. This can not only lead to systemic faults being learned during the operation, but also complicates the evaluation of the estimation quality during the test period. For example, unknown faults that occur after the training period can significantly influence the evaluated metric in the test period. Furthermore, high short-term deviations when the AHU changes from one sequence to another, can have a strong negative influence on the evaluated metric. A complementary analysis of the time series is necessary to compare the dynamics of prediction and observation.

Figure 2 displays the normalised mean absolute error (NMAE) for each regression target of each AHU of the MSA building. To enable a uniform presentation of the results, normalisation is done relatively to the range of values of the respective variable during the test period. To prevent outliers from unduly influencing the results, the lower and upper percentiles at 0.5% and 99.5%, respectively, are used. Not all target variables are shown, as these were identified as categorical in pre-processing and thus other metrics apply. For categorical variables the resulting F1 scores are in a range between 92% and 100%, with two exceptions at 88%, indicating high performance.

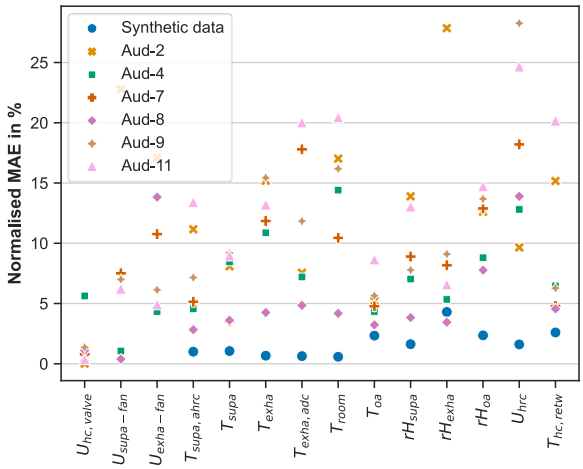

Figure 2: Normalised MAE for the estimators of the AHUs in the MSA building (test period).

It is obvious that the best results are achieved for the synthetic data set, with NMAE-values ranging from 0.4% to 4.3%, resulting in a mean value of 2.4%. This outcome can be attributed to various factors. Firstly, the available data set is larger and enables better model training, as well as a more accurate evaluation on the test data. Moreover, the absence of stochastic influences and unmodeled disturbances contributes to this outcome. It is important to note that the observed system is a closed control loop, and thus, the control behaviour must be learned by the ML model. Furthermore, the implemented control strategies of the AHUs in the MSA building are not well documented and must be derived from the operating data via expert knowledge. However, the control strategy implemented to generate the synthetic data set is much simpler and contains fewer different operating modes. In Figure 3, an exemplary daily trend of the estimated and the observed supply air temperature during the test period of the simulated data set is displayed. As with other target variables, the dynamic trend is accurately estimated, including the heating process in the morning, and set point control during the day.

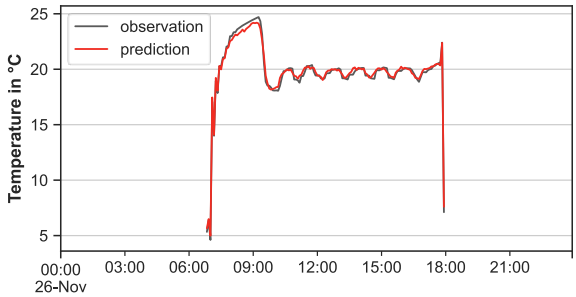

Figure 3: Exemplary daily trend of the observed and estimated supply air temperature in the test period for the simulated data set.

Table 5: Mean of the normalised MAE for the MSA building in training, validation, and test period.

Name of AHU	Training	Validation	Test
Synthetic data	1.40%	1.47%	2.39%
AUD-2	2.80%	4.25%	16.25%
AUD-4	2.88%	4.19%	11.48%
AUD-7	5.30%	7.02%	12.64%
AUD-8	3.44%	5.28%	7.90%
AUD-9	2.78%	6.15%	14.56%
AUD-11	3.38%	5.08%	17.22%

The mean NMAE over all target values are shown in Table 5 for each data set and evaluation period of the MSA building. The results of the test period for the real-life data sets vary in the range of 7,9% to 17.2%. The following analysis of the time series illustrates the deviations exemplary. In Figure 4, the observed supply air temperature of the AUD-8 system on 28.09.2022 is compared with the estimated temperature. The system was in continuous operation and went through different operating states, which is represented by a wide temperature range. Here, the maximum deviation is 0,82 K and the average deviation is 0,05 K. The resulting trend for a summer day (2022-08-24) in Figure 5 shows a noticeably poorer estimation. The dynamics, due to a temporary increased air volume flow, are in principle reproduced, but a continuous offset can be observed. During this period, the system operates in adiabatic cooling mode, which is not well reproduced by the regression model with a peak deviation of 2,4 K.

These examples show that the MLP neural networks can reproduce the dynamic behaviour of the target variables, but not all operating states are captured correctly. Figure 6 and Figure 7 show the proportions of operating modes contained in the respective data sets of two AHUs. The modes are identified by rules. For AHU AUD-8 (Figure 6), the cooling mode is clearly underrepresented and thus not well learned. Comparing the periods for each mode, AHU AUD-8 has very balanced ratios, whereas AUD-7 (Figure 7) shows a clear shift. Especially for the mode classes “return air”, “heating”, and “cooling”, with clearly different proportions in the training and test data set. This explains the very different results for the NMAE value in the respective periods, which are significantly influenced by the class distribution.

In contrast to the MSA building, a longer period of operating data is available for the PSHN building. This allows longer training, validation, and test periods. Additionally, the AHUs are equipped with basic components and the implemented control strategy is less complex. This results in a more balanced data sets with respect to operating modes. Consequently, the estimates for the PSHN building are significantly better than those for the MSA building. The average NMAE values in the test period are between 5.1% and 6.8% and thus close to the results of the training data set (2,6%-4%) (Table 6).

Nevertheless, it can be seen in Figure 8 that individual estimates of the target variables, such as exhaust air temperature and return water temperatures of the heating coil, perform less well. Here, interaction with other components of the HVAC system becomes apparent. Changed setpoint parameters for room air (radiators) and supply water temperature of central heating lead to different operating characteristics compared to the training period. Also noticeable are high NMEA values for the volume flows of the AHU Spo-Hall, where a change in the volume flow setpoint can be observed.

Table 6: Mean of the normalised MAE for the PSHN building in training, validation, and test period

Name of AHU	Training	Validation	Test
Zone-A	2.84%	6.15%	5.34%
Zone-B	2.88%	5.49%	6.78%
Zone-C	2.88%	5.01%	5.06%
Zone-D	3.45%	5.96%	5.74%
Asse-Hall	2.64%	5.57%	5.08%
Spo-Hall	3.98%	5.73%	5.15%

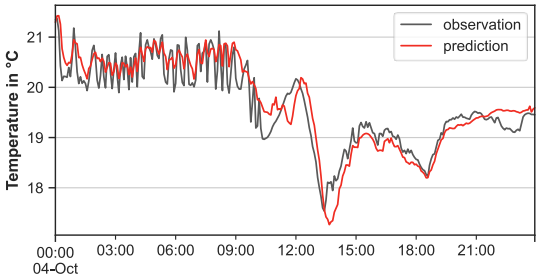

Figure 4: Exemplary daily trend (2022-10-04) of the observed and estimated supply air temperature in the test period for AHU AUD-8 in the MSA building.

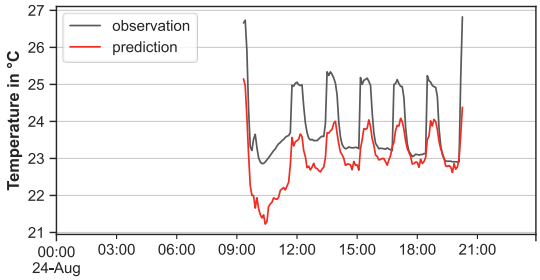

Figure 5: Exemplary daily trend (2022-08-24) of the observed and estimated supply air temperature in the test period for AHU AUD-8 in the MSA building.

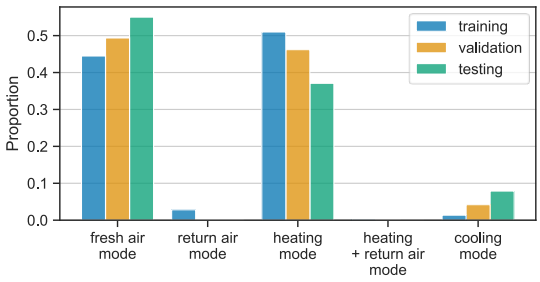

Figure 6: Proportions (range: 0-1) of the operating modes in the datasets of AHU AUD-8 (MSA building).

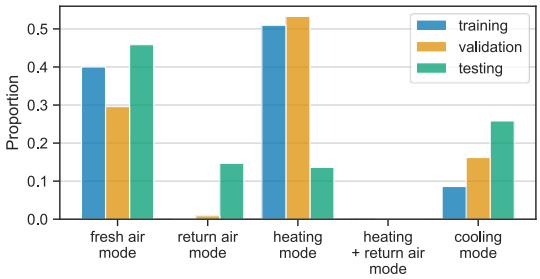

Figure 7: Proportions (range: 0-1) of the operating modes in the datasets of AHU AUD-7 (MSA building).

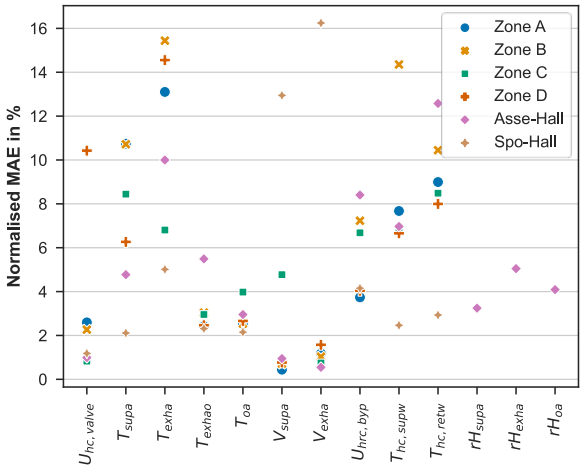

Figure 8: Normalised MAE for the estimators of the AHUs in the PSHN building

Performance of fault detection

The presented FD process capability for fault detection is evaluated using the synthetic data set with known fault conditions and compared to the results of a rule-based fault detection.

Figure 9 shows the results of the residuals-based FD over a period of two years in a confusion matrix. Fault state is detected when the L^2 norm exceeds two times the standard deviation in the training period. 71.6% of the data is classified as true negative and 23.3% as true positive. Only 1.5% of the data cause false alarms and 3.7% of the data are incorrectly classified as fault-free. It should be noted that only periods with fans in operation are evaluated.

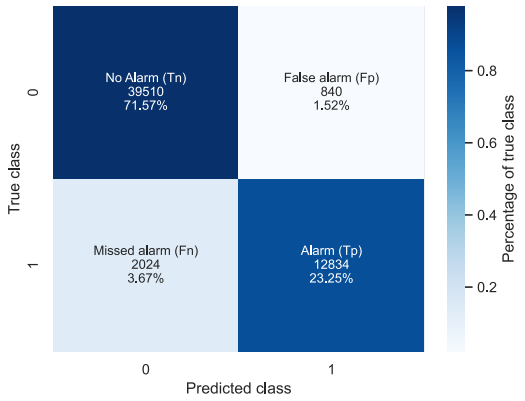

Figure 9: Confusion matrix for the results of FD based on residual generation method. The colour indicates the percentage of the true class (row).

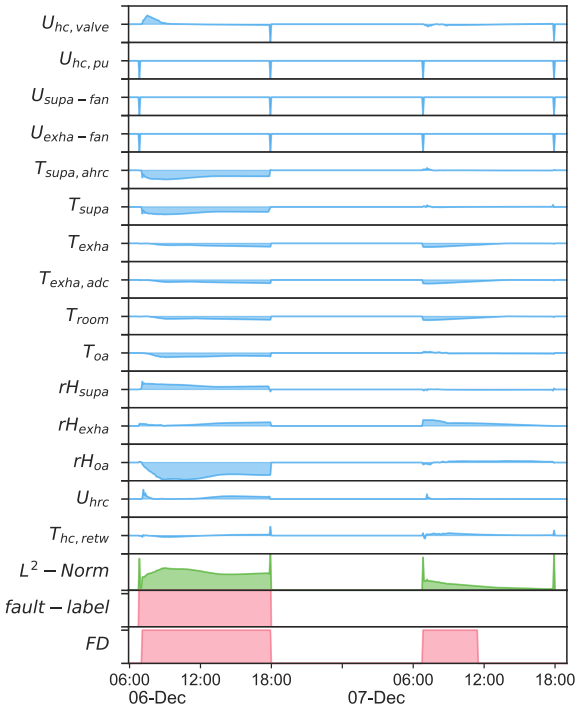

Figure 10: Exemplary trend of the residual vector on the 6th and 7th of Dec. for the synthetic data set. "Fault label" signals the true and "FD" the detected fault state.

Table 7: Metrics for rule- and residuals-based FD on simulated data with known fault states.

Metric	Rule-based	Residuals-based
Classification rate	95.73%	94.81%
Misclassification rate	4.27%	5.19%
False alarm rate	0.01%	2.08%
Fault detection rate	84.17%	86.38%

More insight enables Figure 10. It shows the normalised residual vector, the L2 norm, the true fault condition, and the detected fault status over a period of two days. The fault condition shown is a heat recovery malfunction from 6:00 to 18:00 on the 6th of December. The fault is detected correctly. However, since the zone is undersupplied during the fault period, untypical operating conditions still prevail due to the time constant of the thermal zone. Thus, false alarms are identified when the system starts operating in normal operation the following day. This example shows that counted false alarms usually occur after correctly detected faults. On the other hand, fault conditions with negligible effect on the operating behaviour cause small deviations in the residual vector and are not detected. Since technical building systems are not safety-critical systems, this behaviour is considered as robust.

To rate the performance of the residuals-based method, a second rule-based method is implemented. The kind of faults to be detected are well known and a specific, ideal ruleset can be implemented. For the three faults, a total of seven threshold values with partly interacting relationships are adapted in a manual fine-tuning process to minimise false alarms. This process is expected to be much more complex for real data sets. Additionally, the number of thresholds and thus complexity increases with the number of rules. This is not the case for the proposed residual generating approach.

The metrics in Formulas (9) to (12) are evaluated to assess the accuracy of FD. T_p is the number of true positive, F_p of false positive, T_n of true negative, and F_n of false negative results for n timesteps.

$$Classification\ rate = (T_p + T_n)/n \quad (9)$$

$$Missclassification\ rate = (F_p + F_n)/n \quad (10)$$

$$False\ alarm\ rate = F_p/(F_p + T_n) \quad (11)$$

$$Fault\ detection\ rate = T_p/(T_p + F_n) \quad (12)$$

Table 7 shows the results for both approaches. With the chosen thresholds, the rule-based system with 95.7% as well as the residual generating approach with 94.8% show a very good classification rate. The fault detection rate of the residual generating approach is 86.4%, which is slightly higher than the rule-based system (+2.2%). A false alarm rate of 2.1% is recognisable for the residual generating approach, which can be explained by the system dynamics following high-impact faults, as explained previously. Overall, comparably good results for both approaches are achieved. However, the residual-based approach has the advantage that it requires no complex fine-tuning of numerous threshold values, and it can detect unknown faults.

Conclusion and outlook

In this study, we used a total of twelve real-life datasets from two different AHU types. We have shown that the proposed residual generator, based on MLP networks and Bayesian optimization, is highly transferable, captures the system dynamics, and provides a good estimation of the fault-free behaviour. The process parameters do not need

to be adjusted when applied to a new dataset. Moreover, it is robust against overfitting, temporary outliers, and short-term faults in the training data.

However, the results show that underrepresented operating modes in the training data are not sufficiently learned and lead to poor estimations in the corresponding periods. To improve the estimation in all operation modes, suitable methods for data balancing must be developed and implemented in a next stage. An option is to oversample underrepresented operating modes to train the MLP networks. Additionally, the input data must be continuously checked for changes in the operating conditions compared to the training data. If necessary, the residual generator should be re-trained. This monitoring process should also account for changed set points in other parts of the HVAC system that lie outside the boundary of the observed subsystem. An open question remains regarding the criterion for determining when a trained model is precise enough to be utilized in the residual generating FDD process.

The resulting residual patterns provide the necessary information for both, fault detection and, in a further step, fault diagnosis. In this paper, it is shown that the fault state is determined with a high accuracy over the L^2 norm of the residual vector if applied on a labelled synthetic data set. The threshold used for FD is two times the standard deviation of the L^2 norm in the training period, which accounts model uncertainties for robust fault detection. However, since it is a static threshold, dynamic calculations methods can increase FD sensitivity for low impact faults.

Compared to a rule-based system that can only detect known faults, the proposed approach achieves comparable results: a classification rate of approximately 95% and a fault detection rate of approximately 85%. The residual generating approach offers key advantages, including a high degree of automation and significantly lower setup effort when applied to a new system for observation. Additionally, the residual generation approach can detect unknown faults.

Future work will focus on the application and evaluation of the presented FD approach on labelled real life data sets with known fault conditions. Additionally, the generated residual fault patterns will be evaluated to determine if classification methods can facilitate fault isolation in the fault diagnosis stage of a FDD process.

Acknowledgement

Many thanks to Le Fonds Belval and esp. Mr. G. Spenner for the support of the project with operating data from the Maison du Savoir building. Special thanks to the University of Applied Sciences (HTW) Berlin in the person of Prof. F. Sick, who kindly provide the operating data of the PSHN building.

References

Benndorf, G., Wystrcil, D., Réhault N. (2018). A fault detection system based on two complementary methods and continuous updates. *IFAC-PapersOnLine 51*, 353-358.

Bezyan, B., Zmeureanu, R. (2022). Detection and Diagnosis of Dependent Faults That Trigger False Symptoms of Heating and Mechanical Ventilation Systems Using Combined Machine Learning and Rule-Based Techniques. *Energies 15 (1691)*.

Dietz, S., Sick, F. (2016). Energy-plus primary school Hohen Neuendorf Measurement based evaluation of a hybrid ventilation system. *SBE16 - International Conference on Sustainable Built Environment*. Hamburg (Germany), 488-494, 7th – 11th March 2016.

Ding, S. (2013). Chapter 2 Basic Ideas, Major Issues and Tools in the Observer-Based FDI Framework. In Ding, S. *Model-Based Fault Diagnosis Techniques*. Springer-Verlag. London (England).

Du, Z., Fan, B., Jin, X. et al. (2014). Fault detection and diagnosis for buildings and HVAC systems using combined neural networks and subtractive clustering analysis. *Building and Environment 73*, 1-11.

International Energy Agency IEA (2019). 2019 Global Status report for Buildings and Construction

Katipamula, S. and Brambley, M. (2005). Review article: Methods for fault detection, diagnostics, and prognostics for building systems—A review, part I. *HVAC&R Research*, 3-25.

Lee, W., House, J. and Kyong, N. (2004). Subsystem level fault diagnosis of a building's air-handling unit using general regression neural networks. *Applied Energy 77 (2),* 153–170.

Liao, H., Cai, W., Cheng, F., et al. (2021). An online data-driven fault diagnosis method for air handling units by rule and convolutional neural networks. *Sensors 21 (13)*.

Matetić, I., Štajduhar, I., Wolf, I. et al. (2023). A Review of Data-Driven Approaches and Techniques for Fault Detection and Diagnosis in HVAC Systems. *Sensors 23*, 1-37.

Melgaard, S., Andersen, Kamilla H. et al. (2022). Fault Detection and Diagnosis Encyclopedia for Building Systems: A Systematic Review. *Energies 15 (4366)*.

Mirnaghi, M. and Haghighat, F. (2022). Fault detection and diagnosis of large-scale HVAC systems in buildings using data-driven methods: A comprehensive review. *Energy and Buildings 229 (110249)*.

Réhault, N., Neumann, C., Jakob, D. (2010). Development and application of ongoing commissioning methods and tools for non-residential buildings in German and European research programs. *Proceedings from 6th International Conference on Improving Energy Efficiency in Commercial Buildings.* Frankfurt (Germany), 13-14 April 2010.

Wang, H. and Chen, Y. (2016) A robust fault detection and diagnosis strategy for multiple faults of VAV air handling units., *Energy Build 127*, 442–451.