Delft University of Technology

Delft Center for Systems and Control

Technical report 16-029

Combining knowledge and historical data for system-level fault diagnosis of HVAC systems*

K. Verbert, R. Babuška, and B. De Schutter

If you want to cite this report, please use the following reference instead:
K. Verbert, R. Babuška, and B. De Schutter, "Combining knowledge and historical data for system-level fault diagnosis of HVAC systems," *Engineering Applications of Artificial Intelligence*, vol. 59, pp. 260–273, Mar. 2017.

Delft Center for Systems and Control
Delft University of Technology
Mekelweg 2, 2628 CD Delft
The Netherlands
phone: +31-15-278.24.73 (secretary)
URL: https://www.dcsc.tudelft.nl

Combining Knowledge and Historical Data for System-Level Fault Diagnosis of HVAC systems

K. Verbert[a,*], R. Babuška[a], B. De Schutter[a]

[a]*Delft Center for Systems and Control, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands*

Abstract

Interdependencies among system components and the existence of multiple operating modes present a challenge for fault diagnosis of Heating, Ventilation, and Air Conditioning (HVAC) systems. Reliable and timely diagnosis can only be ensured when it is performed in all operating modes, and at the system level, rather than at the level of the individual components. Nevertheless, almost no HVAC fault diagnosis methods that satisfy these requirements are described in literature. In this paper, we propose a multiple-model approach to system-level HVAC fault diagnosis that takes component interdependencies and multiple operating modes into account. For each operating mode, a distinct Bayesian network (diagnostic model) is defined at the system level. The models are constructed based on knowledge regarding component interdependencies and conservation laws, and based on historical data through the use of virtual sensors. We show that component interdependencies provide useful features for fault diagnosis. Incorporating these features results in better diagnosis results, especially when only a few monitoring signals are available. Simulations demonstrate the performance of the proposed method: faults are timely and correctly diagnosed, provided that the faults result in observable behavior.

Keywords: fault diagnosis, HVAC systems, virtual sensors, Bayesian networks

1. Introduction

Heating, Ventilation, and Air Conditioning (HVAC) systems, widely used in residential and commercial buildings, are responsible for a large part ($20-40\%$) of the worldwide energy consumption [1]. Malfunction or degradation of HVAC system components causes reduced comfort on the one hand, and approximately $15-30\%$ waste of energy on the other hand [2, 3]. Therefore, the development of effective preventive maintenance strategies for HVAC systems is of major importance.

A promising preventive maintenance strategy is condition-based maintenance, which plans the maintenance according to the need indicated by the system condition [4, 5]. An important step within the condition-based maintenance process is the determination of the system condition from the available monitoring signals, hereafter referred to as fault diagnosis [6]. Fault diagnosis of HVAC systems is a challenging task for the following reasons:

1. The HVAC system behavior is difficult to model, as it varies from building to building and it is influenced by uncertain factors, like weather and building use.
2. In general, relatively few variables are measured, especially at the component level. For example, air and water flow rates are rarely available for all components, such as radiators and air handling units.
3. The available measurements are often only crude estimates of the underlying variables, e.g. they are collected by single-point air temperature sensors.
4. (Hierarchical) relationships exist among the different system components [7]. For example, a non-functioning boiler will also affect the working of all radiators and air handling units connected to this boiler. Similarly, the degree to which a radiator fault affects the room temperature depends, among other factors, on the availability and capacity of other radiators in the room.
5. Environmental variations and users settings (e.g. day and night schedules) require that HVAC systems operate in different modes. For example, during the day, both the refresh rate and the supply air temperature are controlled, while during the night only the refresh rate is controlled. Each of the operating modes may require a different diagnostic model.

Although research has been devoted to fault diagnosis for HVAC systems [7–15], almost no attention has been paid to component interdependencies and to the consequences of multiple operating modes. Most papers focus on specific methods (e.g. principal component analysis [10, 11], Bayesian network approaches [16, 17] clustering techniques [15], neural networks [12], fuzzy systems [9, 13], or support vector machines [8, 11, 14]) for the fault di-

*Corresponding author

Email addresses: k.a.j.verbert@tudelft.nl (K. Verbert), r.babuska@tudelft.nl (R. Babuška), b.deschutter@tudelft.nl (B. De Schutter)

agnosis of one specific HVAC component. For example, the authors of [15] propose a model-based diagnosis approach for commercial heat pumps; in [8, 9, 12–14, 17] different diagnosis strategies for the fault diagnosis of an air handling unit have been proposed; [11, 16] specifically focus on the fault diagnosis of the chiller plant; and the authors of [10] present a strategy based on the principal component analysis to detect and diagnose sensor faults in typical air-handling units. In [7] fault diagnosis is considered at the system level taking component interdependence into account. However, the proposed diagnostic model is captured by a rule-based system, which cannot easily be modified to changing situations and other building configurations and which does not take uncertainty into account.

Fault diagnosis methods that do not take both component interdependencies and changing operating modes into account, will not result in adequate fault diagnosis in practice. To ensure correct and timely diagnosis the problem characteristics should explicitly be taken into account in the formulation of the diagnostic model, and that is what we do. More specifically, we propose a multiple-model approach to system-level fault diagnosis in HVAC systems that[1]:

1. takes the interdependencies among the different HVAC components into account; and
2. can easily adapt to changing operation conditions and different building configurations.

Each model is captured by a Bayesian network, which is an intuitive and transparent model for reasoning under uncertainty that can easily adapt to varying operation conditions and different building configurations [18–21]. Bayesian networks have already shown to be an effective reasoning tool for a variety of diagnostic application (see e.g. [16, 22–24]). We construct the Bayesian networks based on both knowledge regarding *component interdependencies* and *conservation laws*, and based on historical data through the use of *virtual sensors*. This way, advantage is taken of the available knowledge and data, while keeping the reasoning transparent. Moreover, we prefer a combined knowledge and data-based approach over a learning-based approach because: 1. the amount of historical data required by learning approaches is not yet available for most buildings; 2. practice calls for an understandable and intuitive decision support system; and 3. the knowledge proposed can support future learning approaches.

The remainder of this paper consists of three parts: In the first part (Sections 2 till Section 4) we propose our multiple-model approach to fault diagnosis in HVAC systems. Next in Sections 5 and 6, we present two case studies involving a simple HVAC-controlled building to demonstrate and evaluate the proposed method. Afterwards, in Section 7 we discuss the generalization to other building configurations.

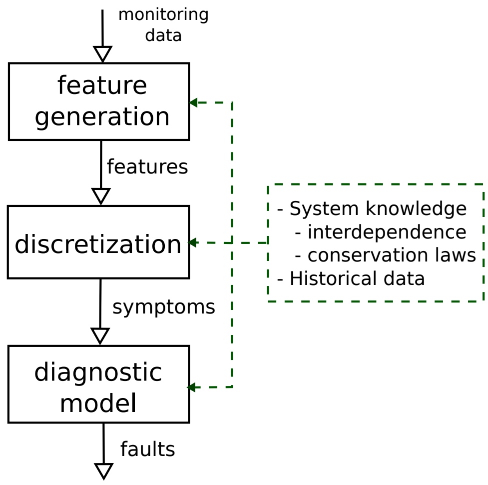


Figure 1: Overview model-based fault diagnosis.

2. Overview of the proposed diagnosis approach

Figure 1 gives a schematic overview of model-based fault diagnosis. First, characteristic features are extracted from the monitoring variables. Next, the continuous-valued features are mapped to discrete-valued symptoms. Finally, based on the symptoms the presence and type of faults is inferred by using the diagnostic model[2].

In our approach, component interdependencies are taken into account by performing diagnosis at the system level (instead of the component level) and by exploiting knowledge regarding component interdependencies in defining the mappings from monitoring data to features, from features to symptoms, and from symptoms to faults (see Figure 1). Because the relations between faults and symptoms are uncertain and may differ for different operating modes, an appropriate diagnostic model is defined for each operating mode and captured by a Bayesian network. Finally, it is ensured that the method is applicable to a wide range of building configurations by exploiting system knowledge that is applicable to all kinds of building configurations (e.g. conservation laws) in defining the different mappings.

3. Elaboration of the diagnosis approach

This section elaborates in more detail on the construction of the Bayesian network, i.e. the diagnostic model. A Bayesian network for a set of variables $\mathbf{X} = \{X_1, ..., X_n\}$ consists of two components [26]:

[1] Challenges 1−3 as outlined before are implicitly accounted for by assuming the availability of only a realistic set of monitored variables, by including noise and measurement error in the simulation model, and by using Bayesian networks to handle the associated uncertainty.

[2] A diagnostic model is a set of static or dynamic relations that link specific input variables – the symptoms – to specific output variables – the faults [25].

1. Network structure N encoding a set of conditional independence assertions about the variables in $\mathbf{X}$;
2. A set $\mathbf{P}$ of local probability distributions associated with each variable.

The network structure N is a directed acyclic graph, with the nodes in one-to-one correspondence to the variables in $\mathbf{X}$ and the edges representing direct dependencies. For more background information on Bayesian networks, see e.g. [26–28]

The construction of a Bayesian network for fault diagnosis consists of the determination of:

1. The network nodes, which can be divided into:
   (a) observable nodes, representing the symptoms;
   (b) unobservable nodes, representing the faults;
2. The probabilistic relations between the nodes.

Because the exact set of symptoms and the relations between symptoms and faults will differ from building to building, we do not propose symptoms here, but we introduce important information sources, namely component interdependencies and conservation laws, and discuss how they can be used for feature extraction and symptom generation. Later, in the case studies in Section 6, symptoms are derived based on these information sources.

3.1. Component interdependencies

In general, an HVAC system can be represented in a hierarchical way as shown in Figure 2. At the top there are the boilers, which provide the air handling units (AHUs) and radiators with hot water. These devices in turn transfer the energy of the hot water to the conditioned zones (radiators) and regulate and circulate the zone air (AHUs). The different components interact in various ways with each other. For the purpose of fault diagnosis, we made a distinction between:

1. *Hierarchical dependencies:* The functioning of a component depends on the proper functioning of higher-level components. For example, when a boiler is not able to heat the water to the desired temperature also the connected radiators and AHUs cannot fulfill their function. When the AHU is not able to adequately condition the air, the connected AHU outlets fail to supply the zone with the desired air.
2. *Compensation by same-level components:* The effect of a non-functioning component can be compensated for by another component fulfilling a similar function. For example, a non-functioning radiator can be compensated for by another radiator in the same zone provided that its capacity is sufficient.

Although the presence of these interdependencies complicates the diagnosis in the sense that the diagnosis cannot be carried out for all components individually, the interdependencies are valuable in the sense that they can serve as characteristic features. Because the interdependencies vary for different faults, their values provide information regarding the fault present. For instance, a boiler fault is probably observed in multiple components or zones, whereas a radiator fault is only locally observed. In this context, an exemplary symptom of a fault in boiler A is "all activate components connected to boiler A are malfunctioning".

3.2. Conservation laws

Both mass and energy balances apply to the HVAC system. Mass balances can be defined for the water flow in the hot water circuit. Energy balances can be defined for each HVAC component where energy is exchanged, e.g. boilers, radiators, and AHUs, and for each conditioned zone. An overview of applicable energy and mass balances can be found in Appendix A.

Energy and mass balances are a useful information source for the formulation of diagnostic features. In the case of a fault, the internal relations between variables or between variables and measurements change. These changes can be detected by verifying internal system relations, including conservation laws. For example, when the measurements do not satisfy the applicable mass balance for the hot water circuit, this could indicate e.g. a leak in the duct work or a sensor fault. In case study 2 in Section 6.3, two features are defined based on, among other things, knowledge regarding mass and energy balances.

3.3. Virtual sensors

Sometimes, the available knowledge is not sufficiently detailed to define the precise relations between features and faults. Consider e.g. that it is known that, in the absence of a particular system fault f_j (i.e. $F_j = 0$), the variable y can be modeled as an unknown function g_1 of variables x_1 and x_2; However, when fault f_j is present, the variable y no longer depends on both x_1 and x_2, but depends only on x_2, i.e.:

$$y = \begin{cases} g_1(x_1, x_2) & \text{if } F_j = 0 \\ g_2(x_2) & \text{if } F_j = 1 \end{cases} \tag{1}$$

From this knowledge, it follows that the symptom "y does not depend on x_1" is characteristic for fault f_j. However, the value of this symptom cannot be assessed based on just this knowledge and instantaneous values of x_1, x_2, and y.

When the available system knowledge is not sufficient to design the diagnostic model, historical data and virtual sensors can be used to complement the available system knowledge, e.g. to find the mapping g_1 in (1). *Virtual sensors* [29, 30] estimate system quantities by using mathematical models, which in turn make use of other physical sensor readings to calculate the estimate. Virtual sensors can be used in the following situations:

1. Absence of a physical sensor, e.g. because the desired quantity cannot be measured or a physical sensor is too slow or costly.
2. As a backup of a physical sensor, i.e. to introduce analytic redundancy. A significant difference between the real sensor and the virtual sensor indicates that one of the two is faulty.

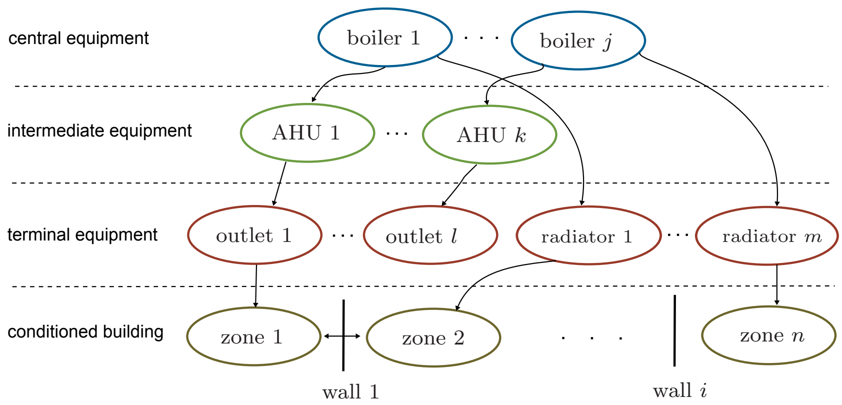


Figure 2: An illustration of the (hierarchical) dependencies among the HVAC components in a building.

3. To estimate the behavior of a system variable corresponding to a specific type of system behavior, e.g. healthy behavior. In this case, the virtual sensor is trained using data corresponding to the considered system behavior and a significant difference between the actual sensor reading and the virtual sensor output indicates that the system does not behave according to the considered behavior.

In the case studies in Section 6, a virtual sensor covering situation 3 is constructed and in Section 7, examples are provided where situation 1 applies.

The design of a virtual sensor essentially consists of three steps:

Step 1: The choice for the quantity to be estimated, i.e. which variables are valuable features for diagnosis.

Step 2: The selection of available sensor measurements that are relevant to estimate these quantities

Step 3: The choice for the method to capture the relation between the quantity of interest and the relevant sensor measurements, e.g. first principles or data-based approaches.

In this work, the main focus is on the first two steps. For the third step, a standard data-based approach from literature, nearest neighbor regression [31], can be used.

4. Fault diagnosis strategy

4.1. Construction of the diagnostic model

Procedure 1 describes the construction of the diagnostic model, in the form of a set of Bayesian networks. In line 1, the system faults f_1 till f_n are determined, e.g. based on expert knowledge. Next, in lines $2-4$, a binary node F_i is assigned to each system fault f_i. Note that a binary node is used for each of the faults to easily handle multiple fault scenarios. Next, in line 5, an appropriate symptom set is determined based on knowledge and data regarding component interdependencies and conservation laws. Subsequently, a node S_j is assigned to each of the symptoms (lines $6-8$). Next, the different operating modes are determined (line 9). For each of them, the relationships between the system faults and the symptoms are defined (i.e. the corresponding network is built) (lines $11-13$).

4.2. Diagnostic inference

For online fault diagnosis, we use the recursive Bayesian estimation scheme as shown in Figure 3, where k denotes a discrete time step and q is the shift operator. In the filtering step, the posterior probability $P(F(k))$ of a fault is determined based on the evidence $S(k)$ and the prior probability $P(\hat{F}(k))$. Based on the outcome of the filtering step, a one-step-ahead prediction $P(\hat{F}(k+1))$ of the fault probability at the next time step is made, which serves as prior for the next filtering step.

In this work, we assume faults to be binary variables, i.e. a fault is either absent or present. In this case, the fault probability at the next time step can only be estimated based on statistical information regarding fault occurrence rates. Since we do not have an accurate predictive model, we assume we assume F to be static, i.e. $P(\hat{F}(k+1)) = P(F(k))$. Now the problem reduces to recursively applying Bayes rule with as prior the previous posterior and as evidence the observations $S(k)$, i.e. we omit the prediction step (see Figure 3(b)).

Please note that in the case that gradually developing faults are considered, the prediction step becomes of interest. In this case, prior knowledge of fault evolution can be combined with observed data.

The recursive diagnosis approach is summarized in Procedure 2. As input it uses the set of Bayesian networks defined in Procedure 1. At each diagnosis instant, first, the actual operating mode is determined (from schedules or measured quantities) (line 3) and next, the corresponding Bayesian network is selected (line 4). Then, based on new evidence e, i.e. observations of the symptoms, the Bayesian network is updated to obtain the posterior fault probabilities (lines $5-7$), which serve as prior probabilities at the next diagnosis instant.

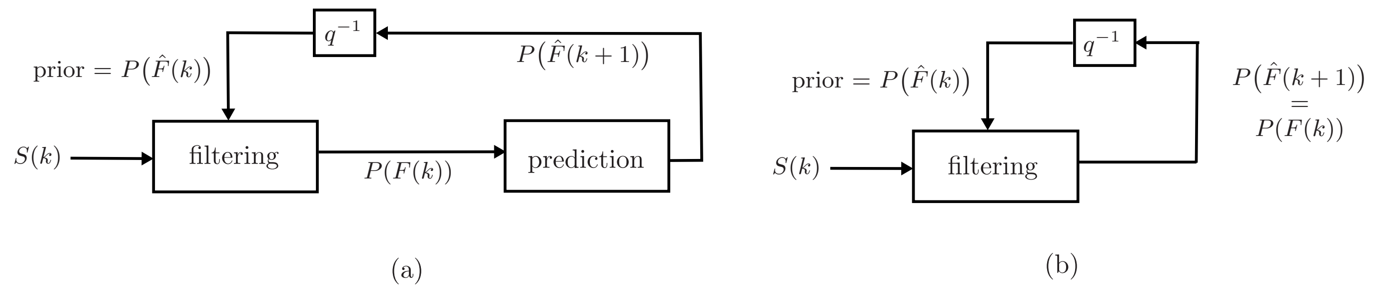


Figure 3: Bayesian fault diagnosis scheme, with evidence $S(k)$ the observations at time k, $F(k)$ the fault variable at time k, and q the shift operator: (a) the full scheme; (b) the simplified scheme adopted in this paper.

Procedure 1 Model construction

Input: Expert knowledge, historical data

Determine possible system faults f_1 till f_n
for $i = 1, ..., n$ **do**
  Define binary node F_i
end for
Determine symptoms S_1 till S_m based on expert knowledge and data
for $j = 1, ..., m$ **do**
  Define discrete-valued node S_j
end for
Determine the system's operating modes 1 till ℓ
for $h = 1, ..., \ell$ **do**
  Determine N_h, which is the network structure defining the relations between the symptoms S_1 till S_m and the fault variables F_1 till F_n, in operating mode h
  Determine $\mathbf{P}_h$, which is the set of local probability functions associated with each node in N_h
end for

Output: Bayesian network $(N_h, \mathbf{P}_h)$ for each operating mode h

Procedure 2 Fault diagnosis

Input: Bayesian network $(N_h, \mathbf{P}_h)$ for each operating mode h, diagnosis instants τ_1 till τ_k

$\mathcal{E} = \{\}$
for $\kappa = 1, ..., k$ **do**
  Determine actual operating mode a at τ_κ
  Select corresponding network $(N_a, \mathbf{P}_a)$
  Store new evidences regarding the symptoms in variable e
  $\mathcal{E} \leftarrow \mathcal{E} \cup \{e\}$
  Update probabilities regarding the faults F_1 till F_n

$$\big(\Pr(F_1|\mathcal{E}), ..., \Pr(F_n|\mathcal{E})\big) = \text{inference}(N_a, \mathbf{P}_a, e)$$
$$\big(\Pr(F_1), ..., \Pr(F_n)\big) \leftarrow \Pr(F_1|\mathcal{E}), ..., \Pr(F_n|\mathcal{E})$$

  with inference$(\cdot)$ the Bayesian inference algorithm
end for

Output: Conditional probability distributions of F_1 till F_n given $\mathcal{E}$

5. HVAC system description

Figure 4 gives an overview of the HVAC configuration considered in this work. The main components are:

1. The zone to be conditioned.
2. HVAC plants, i.e. the equipment installed to control the zone climate.
   (a) Boiler;
   (b) Pump;
   (c) Radiator;
   (d) Air handling unit (AHU).

For the proper understanding of the case studies, some basic knowledge of the AHU and the available monitoring variables is needed.

5.1. Air handling unit

Figure 5 gives an overview of the considered AHU. In the mixing chambers, outdoor air is mixed with air that returned from the zone. The composition of the mixed air is controlled by the positions of three dampers regulating the amount of outdoor air entering the system, the amount of air exhausted from the system, and the amount of return air from the zone to be recirculated. After the mixing, the mixed air passes through the heating coils to condition the air to the desired temperature. The heating in the coils is regulated by the amount and temperature of the water flowing through the coils. The hot water is delivered by the boiler. The temperature of the hot water through the coils is controlled to approximately 40°C using a three-way mixing valve. The amount of water flowing through the coils is determined by the position of a valve, which is controlled by a thermostat based on the differences between the AHU supply air temperature $T^{\mathrm{a}}_{\mathrm{sa}}$ [3] and its setpoint $T^{\mathrm{a}}_{\mathrm{sa,set}}$. Finally, a supply fan is present to maintain a pressure in the supply duct to guarantee that the mixed air is pushed through the coil and finally distributed through the duct work to the zone.

5.2. Monitoring signals

The following monitoring variables are assumed to be available for diagnosis of the considered building:

- Zone air temperature ($T^{\mathrm{z}}_{\mathrm{a}}$)

[3] We use the superscript to indicate the location the variable refers to (e.g. boiler, AHU, zone) and the subscript to indicate the particular mass or air flow (e.g. return water, mixed air).

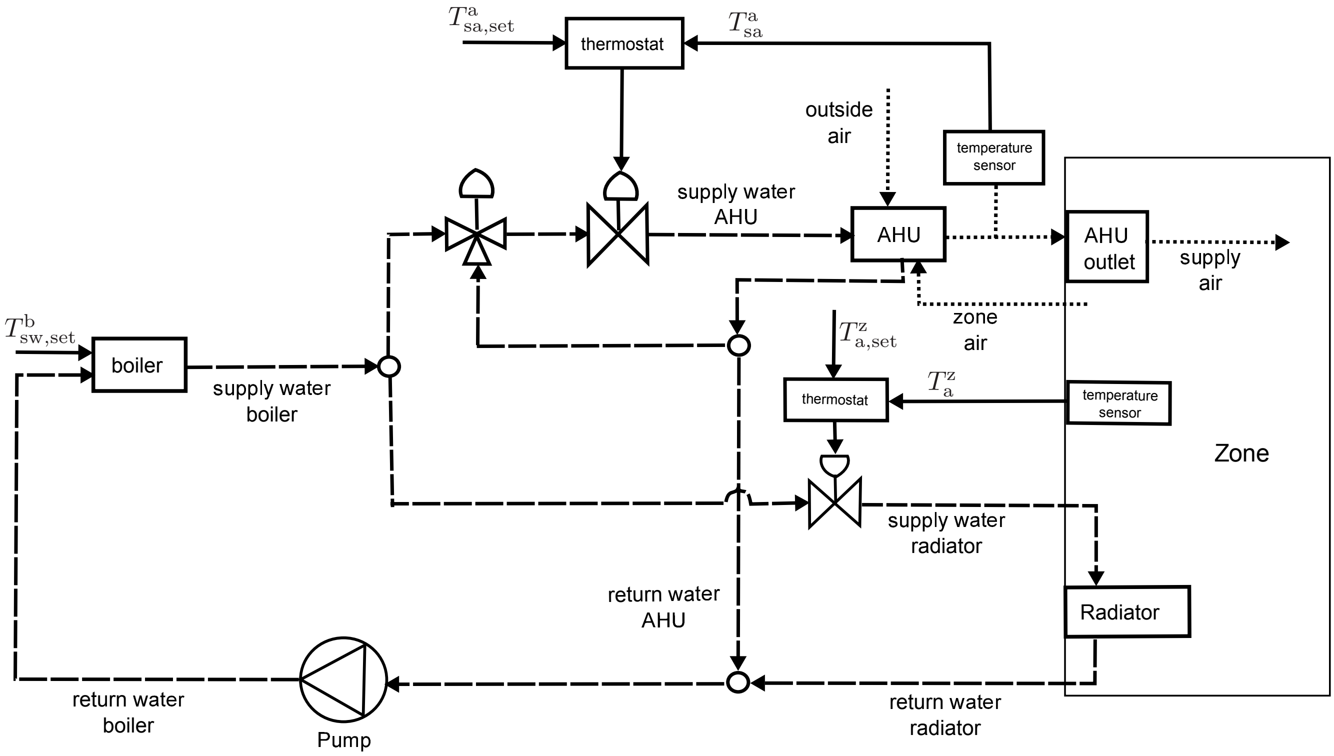


Figure 4: Overview of the considered HVAC system. Dotted lines represent air flows, dashed lines represent mass flows, and solid lines represent signals.

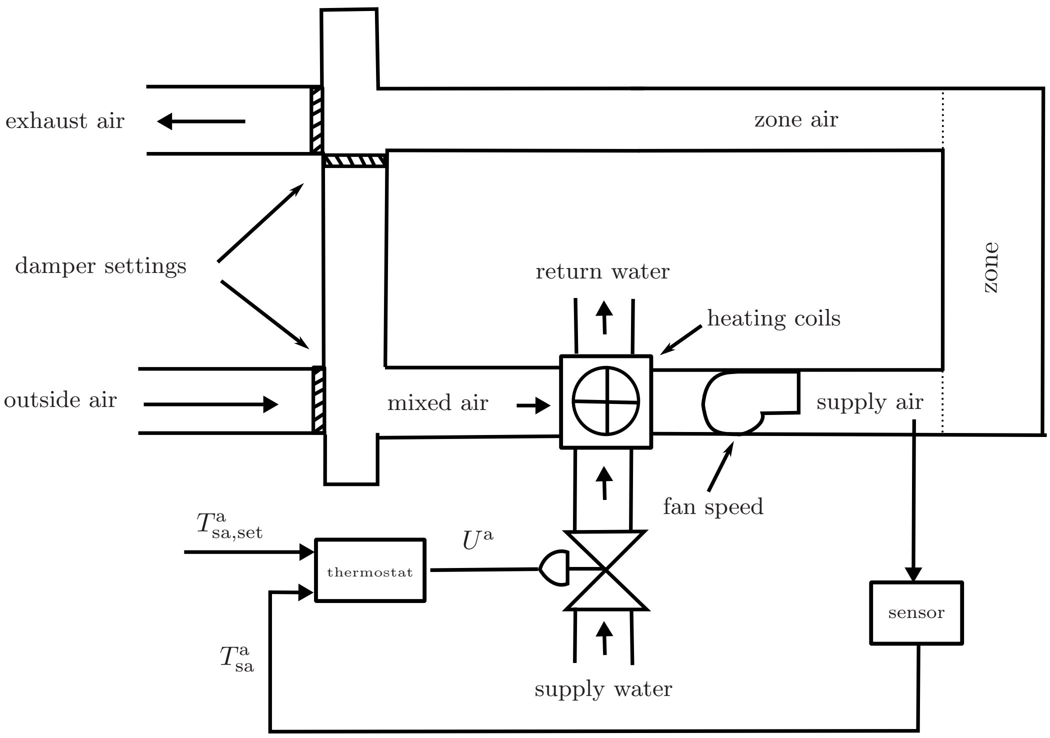


Figure 5: Schematic overview of an AHU.

- Supply air temperature ($T^{\mathrm{a}}_{\mathrm{sa}}$)
- Mixed air temperature ($T^{\mathrm{a}}_{\mathrm{ma}}$)
- Outside air temperature ($T^{\mathrm{o}}_{\mathrm{a}}$)
- Supply water temperature ($T^{\mathrm{b}}_{\mathrm{sw}}$)
- Return water temperature ($T^{\mathrm{b}}_{\mathrm{rw}}$)
- Mass flow through the boiler ($w^{\mathrm{b}}_{\mathrm{sw}}$)
- Control signal to AHU valve (U^{a})
- Control signal to the radiator valve (U^{r})

Furthermore, the zone air temperature setpoint ($T^{\mathrm{z}}_{\mathrm{a,set}}$), supply air temperature setpoint ($T^{\mathrm{a}}_{\mathrm{sa,set}}$), and supply water temperature setpoint ($T^{\mathrm{b}}_{\mathrm{sw,set}}$) are assumed to be known.

6. Fault diagnosis case studies

In this section, the proposed method is illustrated based on two case studies. Case study 1 comprises the fault detection of a stuck AHU heating coil valve and mainly serves to illustrate the problems that occur when neglecting the different operating modes and interdependencies between HVAC components. Case study 2 extends case study 1 in the sense that the possibility of a non-functioning boiler is included. Although this case study is still relatively simple, it clearly illustrates the implications of multiple operation modes and component interdependencies on the fault diagnosis, and how they are handled in the proposed diagnosis approach.

6.1. Simulation model

6.1.1. System modeling

For the purpose of analysis and validation, experts at Honeywell have developed a simulation model of the considered building [32]. The model has been verified using data obtained from real buildings. The model makes a distinction between two sets of variables: temperatures and mass flows. As the pressure dynamics are much faster than the temperature dynamics, the transient behavior of the mass flow rates is neglected, i.e.:

$$w^{\mathrm{a}}_{\mathrm{sw}}(t) = f_{\mathrm{a}}\left(X^{\mathrm{a}}(t), X^{\mathrm{r}}(t)\right) \tag{2}$$
$$w^{\mathrm{r}}_{\mathrm{sw}}(t) = f_{\mathrm{r}}\left(X^{\mathrm{a}}(t), X^{\mathrm{r}}(t)\right) \tag{3}$$

with $w^{\mathrm{a}}_{\mathrm{sw}}$ and $w^{\mathrm{r}}_{\mathrm{sw}}$ the mass flows through the AHU and radiator respectively, and X^{a} and X^{r} the positions of the AHU valve and the radiator valve. For more details on the simulation model, see [32].

6.1.2. Fault modeling

Stuck heating coil valve. A stuck valve stays in the position it was before it got stuck, regardless of the control signal U^{a} sent to the valve by the thermostat. This means that the mass flow through the heating coil remains the same. In the simulation model, a stuck valve is modeled by constraining the mass flow to be constant, i.e.:

$$w^{\mathrm{a}}_{\mathrm{sw}}(t) = w^{\mathrm{a}}_{\mathrm{sw}}(t^{\mathrm{a}}) \quad \forall t \geq t^{\mathrm{a}} \tag{4}$$

with t^{a} the time that the valve stopped functioning.

Non-functioning boiler. When the boiler breaks down, the water returning from the hot water circuit is no longer heated to the supply water temperature setpoint $T^{\mathrm{b}}_{\mathrm{sw,set}}$, i.e. the supply water temperature $T^{\mathrm{b}}_{\mathrm{sw}}$ becomes equal to the return water temperature $T^{\mathrm{b}}_{\mathrm{rw}}$. Therefore, a non-functioning boiler is modeled as follows[4]:

$$T^{\mathrm{b}}_{\mathrm{sw}}(t) = T^{\mathrm{b}}_{\mathrm{rw}}(t^{\mathrm{b}}) \quad \forall t \geq t^{\mathrm{b}} \tag{5}$$

with t^{b} the time that the boiler stopped functioning.

6.1.3. Simulation specifications

1. The daily schedule is defined as:
   - day operation between 04.00 and 18.00 hours;
   - night operation between 18.00 and 04.00 hours.
2. The setpoints of the boiler supply water temperature $T^{\mathrm{b}}_{\mathrm{sw}}$, the AHU supply air temperature $T^{\mathrm{a}}_{\mathrm{sa}}$, and the zone air temperature $T^{\mathrm{z}}_{\mathrm{a}}$ are:

$$T^{\mathrm{b}}_{\mathrm{sw,set}} = \begin{cases} 75 & \text{day operation} \\ 65 & \text{night operation} \end{cases}$$
$$T^{\mathrm{a}}_{\mathrm{sa,set}} = \begin{cases} 20 & \text{day operation} \\ \text{-} & \text{night operation} \end{cases}$$
$$T^{\mathrm{z}}_{\mathrm{a,set}} = \begin{cases} 21 & \text{day operation} \\ 18 & \text{night operation} \end{cases}$$

3. Damper positions are fixed, i.e. the ratio between zone air and outside air is constant (1:4 during the day and 3:7 during the night).
4. Fan speed is fixed, i.e. $w^{\mathrm{a}}_{\mathrm{sa}}$ is constant (0.1kg/s during the day and 0.001kg/s during the night).
5. Detailed weather reports of the winter season are available as input for the simulation.

6.2. Case study 1

Consider the building configuration depicted in Figure 4 and assume that the system is healthy except for a possibly stuck AHU heating coil valve. Our aim is to determine whether or not the valve is stuck. This is a challenging problem because:

[4] Note that in practice there is some delay between the time the boiler stops functioning and the time the supply water temperature becomes equal to the temperature of the return water. We assume this delay to be small and neglect it in the remainder.

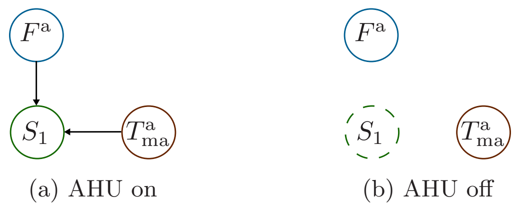


Figure 6: Bayesian network representations of case study 1. During day symptom S_1 is influenced by both an AHU fault and by the mixed air temperature. During night, the AHU is switched off and the relations between F^a, T^a_{ma}, and S_1 no longer hold.

1. The extent to which the fault expresses itself in the measured variables highly depends on the position in which the valve got stuck and on weather conditions;
2. The mass flow through the valve is not measured.

6.2.1. Diagnostic model

Network structure. Given the measurements specified in Section 5.2, an obvious way to detect a stuck heating coil valve is to compare the supply air temperature T^a_{sa} with its setpoint $T^a_{sa,set}$. In the case of a broken valve, a difference between the two temperatures is expected. This knowledge gives rise to define symptom S_1 as:

$$S_1 = \begin{cases} 1 & \text{if } |T^a_{sa} - T^a_{sa,set}| > \epsilon_1 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

with $\epsilon_1 > 0$ a user-defined threshold. The system health is related to symptom S_1 as follows:

If the system is healthy, i.e. $F^a = 0$ **then** likely $S_1 = 0$

If the valve is broken, i.e. $F^a = 1$ **then** likely $S_1 = 1$

with F^a a binary variable indicating whether the AHU valve is healthy ($F^a = 0$) or stuck ($F^a = 1$). Here, "likely" indicates that due to uncertain influences, we are not completely sure about the relations. The degree of uncertainty is expressed in the conditional probability table of S_1, which will be defined later. The relations hold under the assumptions that the system operates in day mode and $T^a_{ma} \leq T^a_{sa,set}$. Because the supply air temperature T^a_{sa} is not controlled during the night, a stuck heating coil valve is only expressed in symptom S_1 during the day. Furthermore, as only heating is present in the considered system, in the summer period when $T^a_{ma} > T^a_{sa,set}$, too high a value of the supply air temperature can be both due to a stuck valve or due to high outside temperatures.

The proposed diagnostic model is graphically represented by the Bayesian networks in Figure 6. Due to the imposed day and night schedule, the system must operate in two modes, which are also reflected in the diagnostic model. As the available simulation data concern the winter season, in which case $T^a_{ma} < T^a_{sa,set}$, node T^a_{ma} is neglected in the remainder.

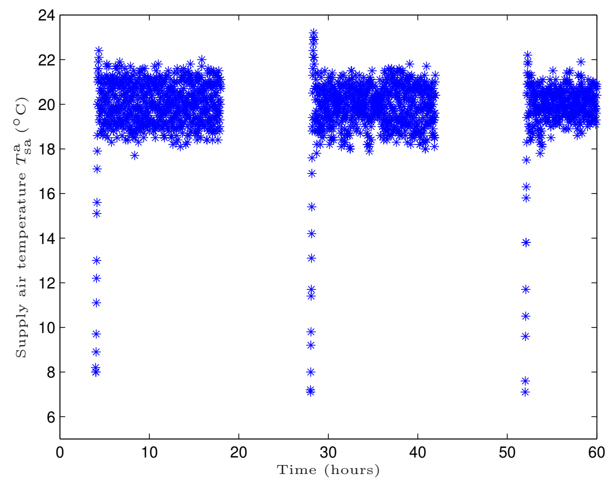


Figure 7: Daily behavior of the supply air temperature. Note that during the night, the AHU supply air temperature is not controlled.

Local probability distributions. To complete the construction of the Bayesian network, the following items need to be determined:

1. the value of ϵ_1;
2. the conditional probability table of S_1;
3. the initial prior probability distribution of F^a.

Determination of ϵ_1 To determine ϵ_1, the nominal variations in T^a_{sa} are considered. Figure 7 shows the behavior of T^a_{sa} on three consecutive days. It can be observed that in the morning, when the system switches to day mode, it takes some time (about half an hour) before the supply air temperature has converged to its desired value $T^a_{sa,set} = 20$°C. After this time, the temperature fluctuates around its desired value. To gain some insight into the degree of fluctuation, in Figure 8 the histogram of $|T^a_{sa} - T^a_{sa,set}|$ containing data of two consecutive months is shown. We tune the value of ϵ_1 such that 99% of the T^a_{sa} values between 04.30 and 18.00 hours are within the interval $[T^a_{sa,set} - \epsilon_1, T^a_{sa,set} + \epsilon_1]$, resulting in

$$\epsilon_1 = 2.5$$

Conditional probability table of S_1 As ϵ_1 is tuned such that in 99% of the healthy cases it holds that $S_1 = 0$, the probability that $S_1 = 1$ given the system is healthy is 1%. To determine the probability that $S_1 = 1$ given a stuck heating coil valve, simulation data from faulty behavior are considered[5]. Actually, the data set used for this must contain measurements corresponding to faults in all different valve positions and for all relevant weather conditions. Figure 9 shows two completely different behaviors

[5]Instead of using (simulation) data, these probabilities can also be directly derived from expert knowledge.

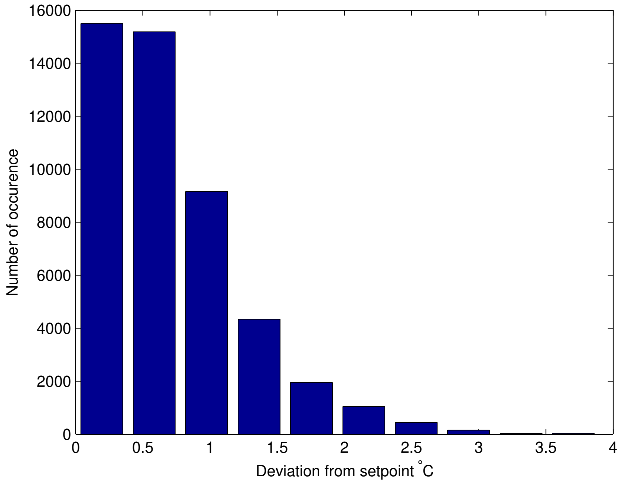


Figure 8: Distribution of $|T^{\mathrm{a}}_{\mathrm{sa}} - T^{\mathrm{a}}_{\mathrm{sa,set}}|$.

Table 1: Conditional probability table of S_1, the values corresponding to $P(S_1|F^{\mathrm{a}})$

	S_1	
F^{a}	0	1
0	0.99	0.01
1	0.24	0.76

of $T^{\mathrm{a}}_{\mathrm{sa}}$ corresponding to a stuck AHU valve. In the first situation, the valve got stuck during night in a cold period, whereas in the second situation, the valve got stuck during day while the outside temperature is increasing. Here, the probability of $S_1 = 1$ given an AHU valve fault ($F^{\mathrm{a}} = 1$) is approximated based on a finite number of randomly chosen fault scenarios. The results are included in Table 1.

Initial prior probability distribution of F^{a} In the first diagnosis step, a user-defined prior $\mathrm{Pr}^0(F^{\mathrm{a}} = 1) = 0.01$ is used. The initial prior probability $\mathrm{Pr}^0(F^{\mathrm{a}} = 1)$ indicates how likely we consider the occurrence of an AHU valve fault before observing the monitoring data. Note that from Bayes' rule, which state that:

$$\mathrm{Pr}(F^{\mathrm{a}}|S_1) = \frac{\mathrm{Pr}(S_1|F^{\mathrm{a}})\,\mathrm{Pr}(F^{\mathrm{a}})}{\sum_{y\in\Theta_{F^{\mathrm{a}}}} \mathrm{Pr}(S_1|y)\,\mathrm{Pr}(y)} \tag{7}$$

with $\Theta_{F^{\mathrm{a}}} = \{0, 1\}$ the domain of F^{a}

it follows that the influence of the initial prior probability distribution on the fault diagnosis is small as the probabilities are recursively updated every minute and the likelihood functions have clearly different values for $F^{\mathrm{a}} = 0$ and $F^{\mathrm{a}} = 1$ (see Table 1).

6.2.2. Fault diagnosis

The proposed approach is demonstrated by means of two simulations. In the first example (see Figure 10), the valve got stuck in a cold period during the night (around time $t = 220$ hours). As a consequence, the air in the AHU is not sufficiently heated during the subsequent day, symptom S_1 becomes equal to one, and shortly afterwards, an AHU fault is detected, i.e. $\hat{F}^{\mathrm{a}} = \mathrm{Pr}(F^{\mathrm{a}} = 1|\mathcal{E}) \approx 1$, where, because of the recursive nature of the Bayesian approach, $\mathcal{E}$ contains all observations of symptom S_1. Besides the correct fault detection around $t = 220$ hours, an AHU fault is incorrectly detected around $t = 160$ hours. This incorrect detection is of a very short duration and a consequence of the way ϵ_1 is tuned. Recall that ϵ_1 is tuned such that in 1% of the healthy cases symptom S_1 is activated. If this happens at several consecutive time instants, this will lead to a false positive detection[6]. In the second example (see Figure 11), the valve got stuck during the day. As the position in which the valve got stuck was quite favorable with respect to the supply air temperature setpoint in the subsequent days, the fault is only detected after four days, i.e. as soon as the effects become observable.

6.2.3. Concluding remarks

Although the diagnostic model defined in Section 6.2.1 turned out to be effective in the sense that in the simulations faults are detected as soon as their effects are observable, diagnosis is not carried out continuously in all operating modes. Specific shortcomings are:

1. Faults cannot be detected during the night;
2. The model is not useful for high mixed-air temperatures;
3. The underlying assumptions are too simplistic, e.g. as only an AHU valve fault is allowed, hierarchical relationships are assumed to be absent.

Therefore, the next section deals with a case study including multiple fault scenarios where the goal is to determine a diagnostic model that it is less sensitive to high values of the mixed air temperature and that allows for fault diagnosis in all operating modes.

6.3. Case study 2

This case study extends the problem discussed in Section 6.2 by including the possibility of a non-functioning boiler. In this case, there are four possible fault scenarios:

1. Healthy system;
2. Stuck heating coil valve:
3. Non-functioning boiler;
4. Both the valve and the boiler are non-functioning.

[6]Remember that we consider a recursive filter in which the poster probabilities serve as prior at the next time step.

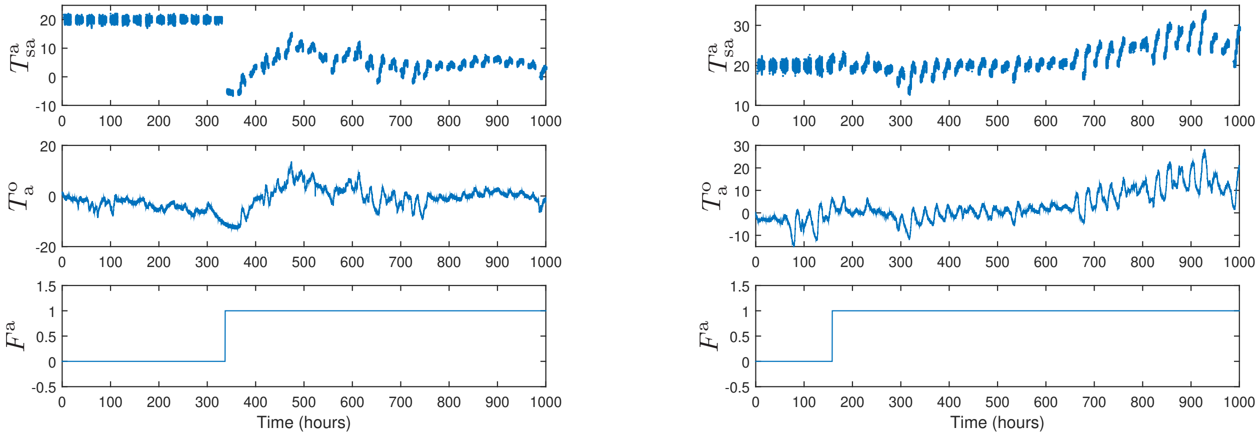


Figure 9: Possible behaviors of $T_{\mathrm{sa}}^{\mathrm{a}}$ corresponding to a stuck heating coil valve. Left: the valve got stuck during the night in a cold period. Right: the valve got stuck during the day while the outside temperature is increasing.

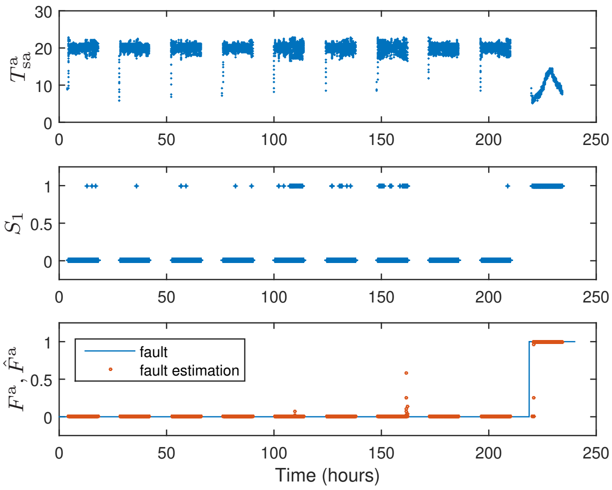


Figure 10: AHU fault diagnosis example 1.

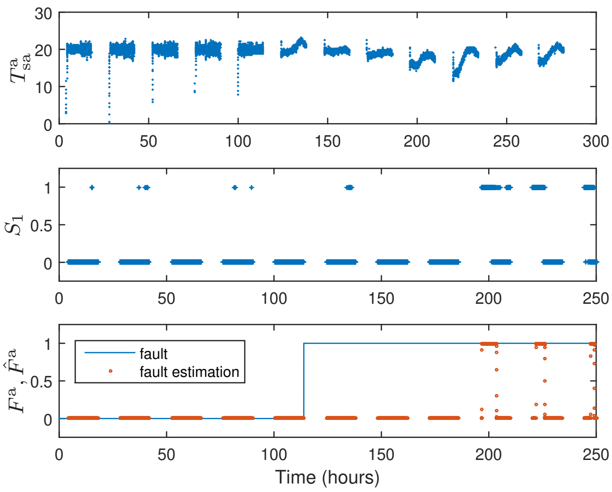


Figure 11: AHU fault diagnosis example 2.

6.3.1. Diagnostic model

Network structure. Besides that the diagnostic model for case study 1 does not support fault diagnosis during the night and is sensitive to high values of the mixed air temperature, the model cannot distinguish between all fault scenarios. If $S_1 = 1$ all scenarios except for scenario 1 are plausible. To make a further distinction between the different fault scenarios possible, symptom S_1 is extended from a binary valued symptom to a three-valued symptom S_1':

$$S_1' = \begin{cases} -1 & \text{if } (T_{\mathrm{sa}}^{\mathrm{a}} - T_{\mathrm{sa,set}}^{\mathrm{a}}) \in (-\infty, -\epsilon_1) \\ 0 & \text{if } (T_{\mathrm{sa}}^{\mathrm{a}} - T_{\mathrm{sa,set}}^{\mathrm{a}}) \in [-\epsilon_1, \epsilon_1] \\ 1 & \text{otherwise} \end{cases} \tag{8}$$

Symptom S_1' relates to the system health as follows:

If $F^{\mathrm{a}} = F^{\mathrm{b}} = 0$ **then** likely $S_1' = 0$

If $F^{\mathrm{a}} = 1$ and $F^{\mathrm{b}} = 0$ **then** likely $S_1' = -1$ or $S_1' = 1$

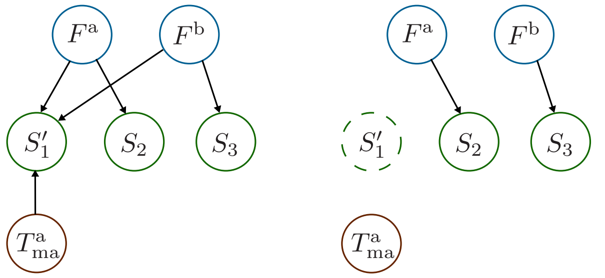
If $F^{\mathrm{b}} = 1$ **then** likely $S_1' = -1$

So, $S_1' = 0$ characterizes a healthy system and $S_1' = 1$ characterizes an AHU valve that got stuck in a too opened position. When $S_1' = -1$, scenarios 2, 3, and 4 are all possible. To improve the diagnostic power and to allow for diagnosis during both the day and the night, two additional symptoms are proposed: S_2 to the verify the proper functioning of the AHU valve and S_3 to verify the proper functioning of the boiler.

To verify whether or not the valve is stuck, the relationships between the mass flow through the boiler $w_{\mathrm{sw}}^{\mathrm{b}}$ and the control signals U^{a} and U^{r} to the AHU valve and the radiator valve respectively are used:

- When $F^{\mathrm{a}} = 0$, the mass flow through the boiler $w_{\mathrm{sw}}^{\mathrm{b}}$ depends both on the control signal to the AHU valve U^{a} and the control signal to the radiator valve U^{r}.

- When $F^{\mathrm{a}} = 1$, the mass flow through the boiler $w_{\mathrm{sw}}^{\mathrm{b}}$ no longer depends on U^{a}, but depends only on U^{r}.



(a) AHU on (b) AHU off

Figure 12: Bayesian network representations of case study 2. During day, symptom S_1' is influenced by both F^{a}, F^{b}, and $T^{\mathrm{a}}_{\mathrm{ma}}$, symptom S_2 is influenced by F^{a}, and symptom S_3 is influenced by F^{b}. During night, when the AHU is switched off, only the relations between F^{a} and S_2 and between F^{b} and S_3 still hold.

This follows from the applicable mass balance (A.1) and equations (2) and (3). Since the relationships among $w^{\mathrm{b}}_{\mathrm{sw}}, U^{\mathrm{a}}$, and U^{r} are not exactly known, we construct a virtual sensor that predicts the mass flow through the boiler $w^{\mathrm{b}}_{\mathrm{sw}}$ based on the AHU and radiator valve control signals U^{a} and U^{r}. The virtual sensor is trained based on healthy data. So, the virtual sensor estimate $\hat{w}^{\mathrm{b}}_{\mathrm{sw}}(U^{\mathrm{a}}, U^{\mathrm{r}})$ will be close to its actual value $w^{\mathrm{b}}_{\mathrm{sw}}$ when the AHU valve functions properly. When the AHU valve is broken, the virtual sensor estimate $\hat{w}^{\mathrm{b}}_{\mathrm{sw}}(U^{\mathrm{a}}, U^{\mathrm{r}})$ likely differs from the measured value $w^{\mathrm{b}}_{\mathrm{sw}}$. This gives symptom S_2 as:

$$S_2 = \begin{cases} 1 & \text{if } |w^{\mathrm{b}}_{\mathrm{sw}} - \hat{w}^{\mathrm{b}}_{\mathrm{sw}}(U^{\mathrm{a}}, U^{\mathrm{r}})| > \epsilon_2 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

Symptom S_2 is linked to the system health as follows:

If $F^{\mathrm{a}} = 0$ **then** likely $S_2 = 0$

If $F^{\mathrm{a}} = 1$ **then** likely $S_2 = 1$

To verify whether the boiler is functioning, a straightforward approach is to compare the boiler supply water temperature $T^{\mathrm{b}}_{\mathrm{sw}}$ with its setpoint $T^{\mathrm{b}}_{\mathrm{sw,set}}$. In case of boiler non-functioning these two values will differ significantly. To this end, symptom S_3 is defined as:

$$S_3 = \begin{cases} 0 & \text{if } (T^{\mathrm{b}}_{\mathrm{sw}} - T^{\mathrm{b}}_{\mathrm{sw,set}}) \in [-\epsilon_3, \infty) \\ 1 & \text{otherwise} \end{cases} \tag{10}$$

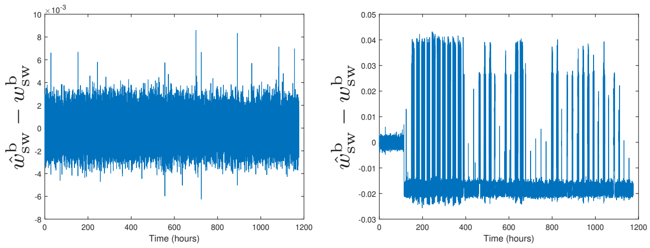
with $\epsilon_3 > 0$, which links to the system health as follows:

If $F^{\mathrm{b}} = 0$ **then** likely $S_3 = 0$

If $F^{\mathrm{b}} = 1$ **then** likely $S_3 = 1$

Considering the symptoms S_1', S_2, and S_3, the diagnostic model for this case is represented by the Bayesian network in Figure 12. A distinction is made between two operating modes: a day mode (AHU on) and a night mode (AHU off). Fault diagnosis can be carried out in both modes.

Similarly as for case study 1, we restrict ourselves to diagnosis in the cold season, i.e. node $T^{\mathrm{a}}_{\mathrm{ma}}$ is disregarded.



(a) $F^{\mathrm{a}} = 0$ (b) $F^{\mathrm{a}} = 1$

Figure 13: Time behavior of $\hat{w}^{\mathrm{b}}_{\mathrm{sw}} - w^{\mathrm{b}}_{\mathrm{sw}}$.

Local probability distributions. Before the network can be used for diagnostic inference, the following items need to be determined:

1. the values of ϵ_1, ϵ_2, and ϵ_3
2. the conditional probability tables of S_1', S_2, and S_3
3. the initial prior probability distributions of F^{a} and F^{b}

Determination of ϵ_1, ϵ_2, and ϵ_3 The value of ϵ_1 is chosen similar as in case study 1 (as the variation of $w^{\mathrm{a}}_{\mathrm{sa}}$ is symmetrical around 20°C, there is no need to make a distinction between positive and negative deviations), i.e.

$$\epsilon_1 = 2.5$$

To determine ϵ_2, the variation in $\hat{w}^{\mathrm{b}}_{\mathrm{sw}} - w^{\mathrm{b}}_{\mathrm{sw}}$ is considered. In Figure 13, time behaviors of $\hat{w}^{\mathrm{b}}_{\mathrm{sw}} - w^{\mathrm{b}}_{\mathrm{sw}}$ are given for both a healthy and a stuck AHU valve. The value of ϵ_2 is chosen such that given $F^{\mathrm{a}} = 0$, it holds that $\Pr(S_2 = 0) = 0.99$. This is the case for

$$\epsilon_2 = 0.003$$

Finally, ϵ_3 is tuned. As the boiler supply water temperature setpoint $T^{\mathrm{b}}_{\mathrm{sw,set}}$ changes at 04.00 hours in the morning and at 18.00 hours in the evening, there is some natural difference between $T^{\mathrm{b}}_{\mathrm{sw}}$ and $T^{\mathrm{b}}_{\mathrm{sw,set}}$ shortly after these times (see Figure 14). Therefore, for fault diagnosis and the determination of ϵ_3, only the time intervals 04.30 till 18.00 hours and 18.30 till 04.00 hours are considered. The value of ϵ_3 is chosen such that given $F^{\mathrm{b}} = 0$, it holds that $\Pr(S_3 = 0) = 0.99$, i.e.:

$$\epsilon_3 = 0.8$$

Conditional probability tables of S_1', S_2, and S_3 The conditional probability tables are defined similarly as in case study 1. The results are given in Tables 2 till 4.

[7] If the boiler is broken the temperature significantly decreases and if the fault holds for some time this probability converges to one.

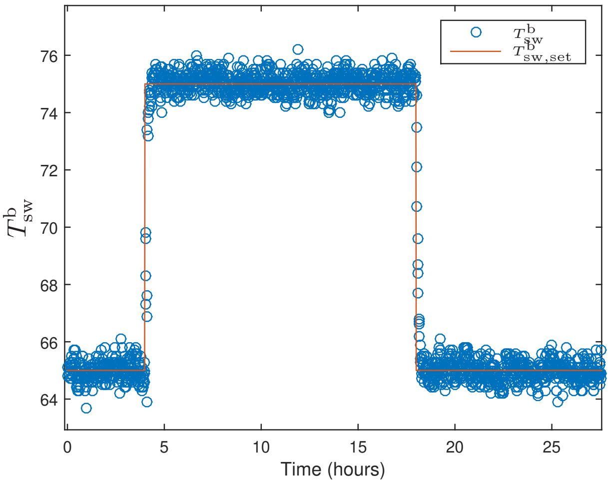


Figure 14: Daily behavior of $T_{\mathrm{sw}}^{\mathrm{b}}$.

Table 2: Conditional probability table of S_1'

F^{a}	F^{b}	S_1: -1	0	1
0	0	0.05	0.99	0.05
1	0	0.47	0.24	0.28
0	1	1^7	0	0
1	1	1	0	0

Table 3: Conditional probability table of S_2

F^{a}	S_2: 0	1
0	0.99	0.01
1	0.11	0.89

Table 4: Conditional probability table of S_3

F^{b}	S_3: 0	1
0	0.99	0.01
1	0	1

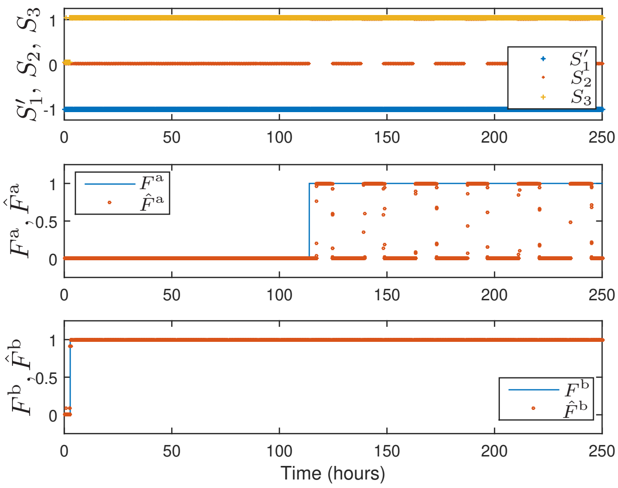


Figure 15: Boiler and AHU fault diagnosis example.

Prior probability distributions of F^{a} and F^{b} The initial prior probability distributions are defined similarly as for case study 1:

$$\Pr^0(F^{\mathrm{a}} = 1) = \Pr^0(F^{\mathrm{b}} = 1) = 0.01$$

Again the effect of the initial priors on the fault diagnosis is small as the likelihood functions have clearly different values for the different fault situations (see Tables 2, 3, and 4).

6.3.2. Fault diagnosis

Consider an example in which the boiler breaks down immediately in the beginning of the simulation and later, at $t = 120$, also the AHU valve gets stuck (see Figure 15). From the simulation results, it follows that the boiler breakdown is clearly expressed in symptoms S_1' and S_3, and that system health is correctly diagnosed till $t \approx 120$ hours, i.e., $\hat{F}^{\mathrm{a}} = \Pr(F^{\mathrm{a}} = 1|\mathcal{E}) \approx 0$, $\hat{F}^{\mathrm{b}} = \Pr(F^{\mathrm{b}} = 1|\mathcal{E}) \approx 1$, where $\mathcal{E}$ contains all observations of S_1', S_2, and S_3. When the AHU gets stuck around $t = 120$ hours also symptom S_2 is activated. Because the position in which the valve got stuck is close to the desired position, symptom S_2 is not continuously activated and the stuck valve is not continuously detected. Even though the fault is not continuously detected, the observed behavior clearly indicates the presence of an AHU valve fault.

6.3.3. Concluding remarks

The proposed diagnostic model overcomes the limitations of the model proposed in case study 1, so that diagnosis is possible in all operating modes, multiple fault situations can be handled, and the model is less sensitive to high values of the mixed air temperature. Furthermore, the diagnostic model has shown to be effective in the considered simulation.

6.4. Alternative symptoms for case study 2

Although the diagnostic model for case study 2 results in good performance, there may exist situations in which other or additional symptoms are required (e.g. in case of an absent or broken supply water temperature sensor). Therefore, we conclude this section with the proposal of two alternative symptoms for case study 2:

1. Find and use the relationship between the supply air temperature $T^{\mathrm{a}}_{\mathrm{sa}}$, the mixed air temperature $T^{\mathrm{a}}_{\mathrm{ma}}$, the supply water temperature $T^{\mathrm{b}}_{\mathrm{sw}}$, and the control signal to the AHU valve U^{a}. Depending on the actual system health, the AHU supply air temperature $T^{\mathrm{a}}_{\mathrm{sa}}$ can be described as a function of:

$$\begin{array}{ll} T^{\mathrm{a}}_{\mathrm{ma}}, U^{\mathrm{a}} & \text{if } F^{\mathrm{a}} = F^{\mathrm{b}} = 0 \\ T^{\mathrm{a}}_{\mathrm{ma}} & \text{if } F^{\mathrm{a}} = 1, F^{\mathrm{b}} = 0 \\ T^{\mathrm{a}}_{\mathrm{ma}}, U^{\mathrm{a}}, T^{\mathrm{b}}_{\mathrm{sw}} & \text{if } F^{\mathrm{a}} = 0, F^{\mathrm{b}} = 1 \\ T^{\mathrm{a}}_{\mathrm{ma}}, T^{\mathrm{b}}_{\mathrm{sw}} & \text{if } F^{\mathrm{a}} = 1, F^{\mathrm{b}} = 1 \end{array} \tag{11}$$

   These relations follow from the energy balance (A.4), the knowledge that the thermal energy of air/water depends on its temperature and volume, and the fact that, for a healthy valve, the mass flow $w^{\mathrm{a}}_{\mathrm{sw}}$ is directly related to the control signal U^{a}. Since the exact relationships are unknown, we use this knowledge to construct two virtual sensors. Multiple virtual sensors are needed since in this case, a distinction between multiple scenarios has to be made[8]. For example, one virtual sensor $\hat{T}^{\mathrm{a}}_{\mathrm{sa}}(T^{\mathrm{a}}_{\mathrm{ma}}, U^{\mathrm{a}})$ is designed to estimate the AHU supply air temperature $T^{\mathrm{a}}_{\mathrm{sa}}$ corresponding to healthy system behavior ($F^{\mathrm{a}} = F^{\mathrm{b}} = 0$) and another one $\tilde{T}^{\mathrm{a}}_{\mathrm{sa}}(T^{\mathrm{a}}_{\mathrm{ma}}, U^{\mathrm{a}}, T^{\mathrm{b}}_{\mathrm{sw}})$ to estimate the behavior of $T^{\mathrm{a}}_{\mathrm{sa}}$ corresponding to a non-functioning boiler ($F^{\mathrm{a}} = 0$, $F^{\mathrm{b}} = 1$). Accordingly, symptom S_{a_1} is defined as (12) and linked to the system health as follows:

   - **If** $F^{\mathrm{a}} = F^{\mathrm{b}} = 0$ **then** likely $S_{\mathrm{a}_1} = 0$
   - **If** $F^{\mathrm{a}} = 0$ and $F^{\mathrm{b}} = 1$ **then** likely $S_{\mathrm{a}_1} = -1$
   - **If** $F^{\mathrm{a}} = 1$ **then** likely $S_{\mathrm{a}_1} = 1$

   A possible drawback of this symptom is that it relies on the availability of historical data of fault situations for designing the virtual sensor (in this case historical data of a non-functioning boiler). However, when a good physical simulator is available, simulated data can also be used to train the virtual sensor.

2. Verify whether other AHUs or radiators connected to the same boiler function properly. This strategy can be used provided that multiple systems (e.g. radiators and AHUs) are connected to the same boiler. In case of a boiler fault, also the connected systems will exhibit aberrant behavior (hierarchical dependencies, see Section 3.1). In the considered building configuration, one radiator is connected to the same boiler as the considered AHU. If this radiator functions properly this indicates that the boiler cannot be broken (provided that radiator heating is required). This knowledge gives rise to defining symptom S_{a_2} as:

$$S_{\mathrm{a}_2} = \begin{cases} 1 & \text{if } T^{\mathrm{z}}_{\mathrm{a}} - T^{\mathrm{z}}_{\mathrm{a,set}} < \epsilon_{\mathrm{a}_2} \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

   which is linked to the system health as follows:

   If $F^{\mathrm{b}} = 0$ **then** likely $S_{\mathrm{a}_2} = 0$

   If $F^{\mathrm{b}} = 1$ **then** likely $S_{\mathrm{a}_2} = 1$

   Note that it is assumed that the radiator functions properly and that this symptom is only useful when radiator heating is required.

Taking the additional symptoms S_{a_1} and S_{a_2} into account the diagnostic model is represented by the Bayesian network in Figure 16. Now, a distinction between four operating modes has to be made. An advantage of this model compared to the original model (see Figure 12) is that, due to its redundancy, fault diagnosis is also possible when one of the symptoms is missing. In addition, the redundancy can be used to detect possible sensor faults.

7. Discussion on generalization

So far, the focus was on one particular HVAC configuration. In practice, each building is different, e.g. it may have another number of zones, different types of separation between the zones, and different HVAC equipment installed to condition the building. Therefore, it is important to consider how the diagnostic model can be extended to other cases.

7.1. Different HVAC equipment

In general, a building (including HVAC system) can be represented as shown in Figure 2. The number of components in each layer and the way the components are connected varies from building to building. These differences influence the diagnostic model. Here, it is shown that even for two slightly different HVAC configurations the diagnostic model may vary. For this purpose, an additional radiator is installed in the building setup considered before (Figure 2). In the original building, a non-functioning radiator, $F^{\mathrm{r}} = 1$, will manifest itself in a too low zone temperature (provided that radiator heating is required). This gives rise to use symptom S_{g_1}, which is defined as:

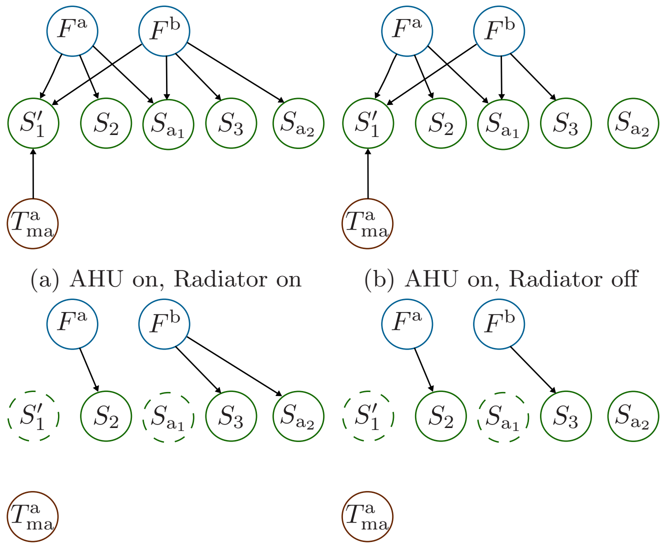
$$S_{\mathrm{g}_1} = \begin{cases} 1 & \text{if } T^{\mathrm{z}}_{\mathrm{a}} - T^{\mathrm{z}}_{\mathrm{a,set}} < -\epsilon_{\mathrm{g}_1} \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

and linked to the system health as:

If $F^{\mathrm{r}} = 0$ **then** likely $S_{\mathrm{g}_1} = 0$

[8]For sake of clarity, we restrict ourselves to two virtual sensors here.

$$S_{a_1} = \begin{cases} -1 & \text{if } \left|T^a_{sa} - \tilde{T}^a_{sa}(T^a_{ma}, U^a, T^b_{sw})\right| < \epsilon_{a_1} \quad \text{and} \quad \left|T^a_{sa} - \tilde{T}^a_{sa}(T^a_{ma}, U^a, T^b_{sw})\right| \le \left|T^a_{sa} - \hat{T}^a_{sa}(T^a_{ma}, U^a)\right| \\ 0 & \text{if } \left|T^a_{sa} - \hat{T}^a_{sa}(T^a_{ma}, U^a)\right| < \epsilon_{a_1} \quad \text{and} \quad \left|T^a_{sa} - \tilde{T}^a_{sa}(T^a_{ma}, U^a, T^b_{sw})\right| > \left|T^a_{sa} - \hat{T}^a_{sa}(T^a_{ma}, U^a)\right| \\ 1 & \text{otherwise} \end{cases} \tag{12}$$



(a) AHU on, Radiator on (b) AHU on, Radiator off

(c) AHU off, Radiator on (d) AHU off, Radiator off

Figure 16: Bayesian network of case study 2 with alternative symptoms taken into account.

If $F^r = 1$ **then** likely $S_{g_1} = 1$

In the new building, this relation does not necessarily hold. A non-functioning radiator may be compensated for by the other radiator, provided that its capacity is sufficient. In this case, a non-functioning radiator needs to be identified in an alternative way, e.g. by verifying whether the radiator control signal U^r is close to control signal expected based on the outside temperature $\hat{U}^r(T^o_a)$. This means that the Bayesian network should be extended with an extra symptom node S_{g_2} connected to F^r, with:

$$S_{g_2} = \begin{cases} 1 & \text{if } |U^r - \hat{U}^r(T^o_a)| > \epsilon_{g_2} \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

with $\hat{U}^r(T^o_a)$ a prediction of U^r based on weather information. Symptom S_{g_2} relates to the system health as:

If $F^r = 0$ **then** likely $S_{g_2} = 0$

If $F^r = 1$ **then** likely $S_{g_2} = 1$

7.2. Different monitoring variables

The symptoms proposed in this work rely on the availability of monitoring data (see Table 5 for an overview of the variables required by each of the proposed symptoms). The set of available monitoring signals however varies from building to building. This means that there may exist situations in which part of the monitoring data required to compute the underlying features is missing. In this case, one of the following strategies can be followed:

Table 5: Variables required by each of the proposed symptoms.

Symptom	Required variables
S'_1	T^a_{sa}, $T^a_{sa,set}$
S_2	w^b_{sw}, U^a, U^r
S_3	T^b_{sw}, $T^b_{sw,set}$
S_{a_1}	T^a_{sa}, T^a_{ma}, U^a, T^b_{sw}
S_{a_2}	U^r, T^z_a, $T^z_{a,set}$

1. definition of alternative symptoms;
2. use of virtual sensors to estimate missing variables.

The first strategy searches for alternative symptoms that can be determined from the available monitoring data and that can replace the missing original symptoms. Consider for example that the control signal to the radiator valve U^r is not measured, meaning that symptom S_2 cannot be defined. In this case, another symptom is needed to identify a stuck AHU heating coil valve. When both the control signal to the AHU valve U^a, i.e. the desired position of the valve, and the actual position of the valve X^a are available, a straightforward alternative symptom S_{g_3} is:

$$S_{g_3} = \begin{cases} 1 & \text{if } |U^a - X^a| > \epsilon_{g_3} \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

which relates to the system health as:

If $F^a = 0$ **then** likely $S_{g_3} = 0$

If $F^a = 1$ **then** likely $S_{g_3} = 1$

In practice, the definition of adequate alternative symptoms is often not so obvious. In this case, strategy 2 becomes of interest, which aims to estimate the missing variable based on the available variables using a virtual sensor. Considering again that U^r is not measured, then symptom S_2 can still be used if U^r can be accurately estimated based on the available data, e.g. by estimating U^r based on the zone air temperature T^z_a and its setpoint $T^z_{a,set}$.

7.3. Different control strategies

The way in which the different temperatures and mass flows in the HVAC systems are controlled influences the diagnostic model. For example, in the case studies considered in Section 6, the fan speed and so the air flow w^a_{sa} through the AHU are fixed. This justifies that for symptom S_{a_1}, only U^a, T^a_{ma}, and T^b_{sw} are used as inputs for the virtual sensor. However, when the fan speed is controlled, a correct implementation of symptom S_{a_1} requires the mass flow rate $w^{\mathrm{a}}_{\mathrm{sa}}$ to be included as input of the virtual sensor. Indeed, when $w^{\mathrm{a}}_{\mathrm{sa}}$ varies over time, there is no fixed relation between $T^{\mathrm{a}}_{\mathrm{sa}}$ and U^{a} and $T^{\mathrm{a}}_{\mathrm{ma}}$ for a healthy system, and no fixed relation between $T^{\mathrm{a}}_{\mathrm{sa}}$ and U^{a}, $T^{\mathrm{a}}_{\mathrm{ma}}$, and $T^{\mathrm{b}}_{\mathrm{sw}}$ in case of a non-functioning boiler. Similarly, in systems where the supply water temperature $T^{\mathrm{a}}_{\mathrm{sw}}$ to the AHU is not controlled to a fixed value, this variable should be included as an input of the virtual sensor.

8. Conclusions

In this work, a model-based Bayesian network approach to fault diagnosis in HVAC systems has been proposed. The diagnostic model was defined using expert knowledge regarding component interdependencies and conservation laws and historical data by the use of virtual sensors. Important properties of the proposed method are: 1. it adequately handles interdependencies between the different components, 2. diagnosis is carried out continuously in all operating modes, and 3. the method is applicable to all kinds of building setups. The importance of these properties and the applicability of the proposed method have been demonstrated based on various case studies. It is concluded that faults are timely and properly diagnosed, even in the case of multiple faults, provided that the fault results in any undesired behavior.

Because a different diagnostic model is required for each building and each operation mode, a lot of time and effort is saved when the diagnostic model can be automatically generated for a class of common buildings and operating modes. In future work, we will therefore work on methods to automate the construction of the diagnostic model. Another direction for future research includes the extension of the method to other diagnostic applications. Indeed, most of the method ingredients, e.g. exploiting component interdependencies, and combining knowledge and data, are applicable to other applications as well. Potential applications include e.g. fault diagnosis of road and railway networks.

Acknowledgment

We thank Ondrej Holub, Jan Berka, Karel Macek, and Henrik Dibowski of the Honeywell Prague Labs for the provision of the simulation data and for the fruitful discussions during our stays at Honeywell Prague Labs.

The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement nr. 324432 (AMBI project). This research is part of STW/ProRail project "Advanced monitoring of intelligent rail infrastructure (ADMIRE)", project 12236, which is supported by ProRail and the Dutch Technology Foundation STW, which is part of the Netherlands Organization for Scientific Research (NWO), and which is partly funded by the Ministry of Economic Affairs.

References

[1] L. Pérez-Lombard, J. Ortiz, C. Pout, A review on buildings energy consumption information, Energy and BFuildings 40 (3) (2008) 394–398.

[2] M. A. Piette, S. K. Kinney, P. Haves, Analysis of an information monitoring and diagnostic system to improve building operations, Energy and Buildings 33 (8) (2001) 783–791.

[3] S. Katipamula, M. R. Brambley, Review article: methods for fault detection, diagnostics, and prognostics for building systemsa review, part i, HVAC&R Research 11 (1) (2005) 3–25.

[4] L. Wang, Modeling and simulation of HVAC faulty operations and performance degradation due to maintenance issues, in: Proceedings of the Asia Conference of International Building Performance Simulation Association, 2014.

[5] R. Yam, P. Tse, L. Li, P. Tu, Intelligent predictive decision support system for condition-based maintenance, The International Journal of Advanced Manufacturing Technology 17 (5) (2001) 383–391.

[6] A. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mechanical Systems and Signal Processing 20 (7) (2006) 1483–1510.

[7] J. Schein, S. T. Bushby, A hierarchical rule-based fault detection and diagnostic method for HVAC systems, HVAC&R Research 12 (1) (2006) 111–125.

[8] J. Liang, R. Du, Model-based fault detection and diagnosis of HVAC systems using support vector machine method, International Journal of Refrigeration 30 (6) (2007) 1104–1114.

[9] A. L. Dexter, D. Ngo, Fault diagnosis in air-conditioning systems: a multi-step fuzzy model-based approach, HVAC&R Research 7 (1) (2001) 83–102.

[10] S. Wang, F. Xiao, AHU sensor fault diagnosis using principal component analysis method, Energy and Buildings 36 (2) (2004) 147–160.

[11] S. M. Namburu, M. S. Azam, J. Luo, K. Choi, K. R. Pattipati, Data-driven modeling, fault diagnosis and optimal sensor selection for HVAC chillers, IEEE Transactions on Automation Science and Engineering, 4 (3) (2007) 469–473.

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

[13] C. Lo, P. Chan, Y. Wong, A. Rad, K. Cheung, Fuzzy-genetic algorithm for automatic fault detection in HVAC systems, Applied Soft Computing 7 (2) (2007) 554–560.

[14] T. Mulumba, A. Afshari, K. Yan, W. Shen, L. K. Norford, Robust model-based fault diagnosis for air handling units, Energy and Buildings 86 (2015) 698–707.

[15] D. Zogg, E. Shafai, H. Geering, Fault diagnosis for heat pumps with parameter identification and clustering, Control Engineering Practice 14 (12) (2006) 1435–1444.

[16] Y. Zhao, F. Xiao, S. Wang, An intelligent chiller fault detection and diagnosis methodology using bayesian belief network, Energy and Buildings 57 (2013) 278 – 288.

[17] F. Xiao, Y. Zhao, J. Wen, S. Wang, Bayesian network based {FDD} strategy for variable air volume terminals, Automation in Construction 41 (2014) 106 – 118.

[18] A. Darwiche, Modeling and Reasoning with Bayesian Networks, Cambridge University Press, 2009.

[19] W. Wiegerinck, H. Kappen, W. Burgers, Bayesian networks for expert systems: Theory and practical applications, in: R. Babuška, F. Groen (Eds.), Interactive Collaborative Information Systems, Vol. 281 of Studies in Computational Intelligence, Springer, 2010, pp. 547–578.

[20] J. Pearl, S. Russel, Bayesian networks, in: M. Arbib (Ed.),

Handbook of Brain Theory and Neural Networks, MIT press, 2001.

[21] H. Boudali, J. B. Dugan, A discrete-time Bayesian network reliability modeling and analysis framework, Reliability Engineering & System Safety 87 (3) (2005) 337–349.

[22] Z. Yongli, H. Limin, L. Jinling, Bayesian networks-based approach for power systems fault diagnosis, IEEE Transactions on Power Delivery 21 (2) (2006) 634–639.

[23] F. Sahin, M. etin Yavuz, Z. Arnavut, nder Uluyol, Fault diagnosis for airplane engines using bayesian networks and distributed particle swarm optimization, Parallel Computing 33 (2) (2007) 124 – 143.

[24] B. Cai, Y. Liu, Q. Fan, Y. Zhang, Z. Liu, S. Yu, R. Ji, Multi-source information fusion based fault diagnosis of ground-source heat pump using bayesian network, Applied Energy 114 (2014) 1 – 9.

[25] R. Isermann, Fault-Diagnosis Applications: Model-Based Condition Monitoring: Actuators, Drives, Machinery, Plants, Sensors, and Fault-Tolerant Systems, Springer Science & Business Media, 2011.

[26] D. Heckerman, A tutorial on learning with Bayesian networks, in: M. Jordan (Ed.), Learning in Graphical Models, Vol. 89, Springer Netherlands, 1998, pp. 301–354.

[27] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann Publishers, 1988.

[28] J. Pearl, Causality; Models, Reasoning, and Inference, Cambridge University Press, 2000.

[29] H. Li, J. E. Braun, Decoupling features and virtual sensors for diagnosis of faults in vapor compression air conditioners, International Journal of Refrigeration 30 (3) (2007) 546–564.

[30] M. Oosterom, R. Babuška, Virtual sensor for fault detection and isolation in flight control systems-fuzzy modeling approach, in: Proceedings of the 39th IEEE Conference on Decision and Control, Vol. 3, 2000, pp. 2645–2650.

[31] N. S. Altman, An introduction to kernel and nearest-neighbor nonparametric regression, The American Statistician 46 (3) (1992) 175–185.

[32] O. Holub, K. Macek, HVAC simulation model for advanced diagnostics, in: Proceedings of the IEEE 8th International Symposium on Intelligent Signal Processing, 2013, pp. 93–96.

Appendix A. Energy and mass balances

For each hot water circuit in the HVAC, the following mass balance applies:

$$w^{\mathrm{b}}_{\mathrm{sw}}(t) = w^{\mathrm{a}_1}_{\mathrm{sw}}(t) + \ldots + w^{\mathrm{a}_{n_\mathrm{a}}}_{\mathrm{sw}}(t) + w^{\mathrm{r}_1}_{\mathrm{sw}}(t) + \ldots + w^{\mathrm{r}_{n_\mathrm{r}}}_{\mathrm{sw}}(t) \tag{A.1}$$

with $w^{\mathrm{b}}_{\mathrm{sw}}(t)$ the mass flow through the boiler at time t, and $w^{\mathrm{a}_1}_{\mathrm{sw}}(t) + \ldots + w^{\mathrm{a}_{n_\mathrm{a}}}_{\mathrm{sw}}(t)$ and $w^{\mathrm{r}_1}_{\mathrm{sw}}(t) + \ldots + w^{\mathrm{r}_{n_\mathrm{r}}}_{\mathrm{sw}}(t)$, the mass flows through the connected AHUs and radiators respectively at time t.

Energy balances can be defined for each component in the HVAC system where energy is exchanged, e.g. the boiler, the radiator, and the AHU. In the boiler, chemical or electrical energy is transformed into thermal energy. The heat generated is used to warm up the water in the hot water circuit. So, the following energy balance holds:

$$E^{\mathrm{b}}_{\mathrm{chem}}(t-\Delta) - E^{\mathrm{b}}_{\mathrm{chem}}(t) = \int_{t-\Delta}^{t} \left(E^{\mathrm{b}}_{\mathrm{sw,thermal}}(\tau) - E^{\mathrm{b}}_{\mathrm{rw,thermal}}(\tau) + E^{\mathrm{b}}_{\mathrm{loss}}(\tau)\right) d\tau \tag{A.2}$$

with $E^{\mathrm{b}}_{\mathrm{chem}}$ the energy in the available fuel, $E^{\mathrm{b}}_{\mathrm{rw,thermal}}$ the thermal energy of the water returning from the hot water circuit, $E^{\mathrm{b}}_{\mathrm{sw,thermal}}$ the energy in the water after it is heated by the boiler, $E^{\mathrm{b}}_{\mathrm{loss}}$ all energy originating from the fuel that is not converted to thermal energy of the water, and Δ a time shift.

In the radiator, part of the thermal energy of the hot water is transferred to the neighboring air of relatively low temperature. The degree of energy exchange depends on the difference between the temperature of the hot water flowing through the radiator and the temperature of the zone air. The following energy balance applies:

$$E^{\mathrm{r}}_{\mathrm{sw,thermal}}(t) - E^{\mathrm{r}}_{\mathrm{rw,thermal}}(t) = Q^{\mathrm{r}}(t) + E^{\mathrm{r}}_{\mathrm{loss}}(t) \tag{A.3}$$

with $E^{\mathrm{r}}_{\mathrm{sw,thermal}}$ and $E^{\mathrm{r}}_{\mathrm{rw,thermal}}$ the thermal energy of the radiator supply and return water respectively, Q^{r} the heat transferred to the zone, and $E^{\mathrm{r}}_{\mathrm{loss}}$ the energy extracted from the water that is not transferred to the zone.

The energy exchange in the AHU is similar to that in the radiator, i.e. thermal energy of the water flowing through the coils is used to increase the thermal energy of the passing air:

$$E^{\mathrm{a}}_{\mathrm{sw,thermal}}(t) - E^{\mathrm{a}}_{\mathrm{rw,thermal}}(t) = E^{\mathrm{a}}_{\mathrm{sa,thermal}}(t) - E^{\mathrm{a}}_{\mathrm{ma,thermal}}(t) + E^{\mathrm{a}}_{\mathrm{loss}}(t) \tag{A.4}$$

with $E^{\mathrm{a}}_{\mathrm{sw,thermal}}$ and $E^{\mathrm{a}}_{\mathrm{rw,thermal}}$ the thermal energy of the AHU return and supply water respectively, $E^{\mathrm{a}}_{\mathrm{sa,thermal}}$ and $E^{\mathrm{a}}_{\mathrm{ma,thermal}}$ the thermal energy of the supply air and the mixed-air respectively, and $E^{\mathrm{a}}_{\mathrm{loss}}$ energy losses. In addition to the energy balances for the HVAC system components, energy balances apply to the zone(s):

$$m_{\mathrm{z}} c_{\mathrm{z}} \dot{T}^{\mathrm{z}}_{\mathrm{a}}(t) = -Q^{\mathrm{z}}(t) + Q^{\mathrm{r}}(t) + Q^{\mathrm{a}}(t) + Q^{\eta}(t) + \sigma(t) \tag{A.5}$$

with $T^{\mathrm{z}}_{\mathrm{a}}$ the zone air temperature, $m_{\mathrm{z}} c_{\mathrm{z}}$ the thermal capacity of the zone, Q^{z} heat losses to the outside/other zones, Q^{r} the heat produced by the radiators, Q^{a} the heat produced by the AHUs, Q^{η} the heat produced by people inside the room, and σ modeling and process noise.