



HAL
open science

Integrated building fault detection and diagnosis using data modeling and Bayesian networks

Tianyun Gao, Sylvain Marié, Patrick Béguery, Simon Thebault, Stéphane Lecoecuche

► **To cite this version:**

Tianyun Gao, Sylvain Marié, Patrick Béguery, Simon Thebault, Stéphane Lecoecuche. Integrated building fault detection and diagnosis using data modeling and Bayesian networks. *Energy and Buildings*, 2024, 306, pp.113889. 10.1016/j.enbuild.2024.113889 . hal-04423161

HAL Id: hal-04423161

<https://hal.science/hal-04423161>

Submitted on 30 Jan 2024

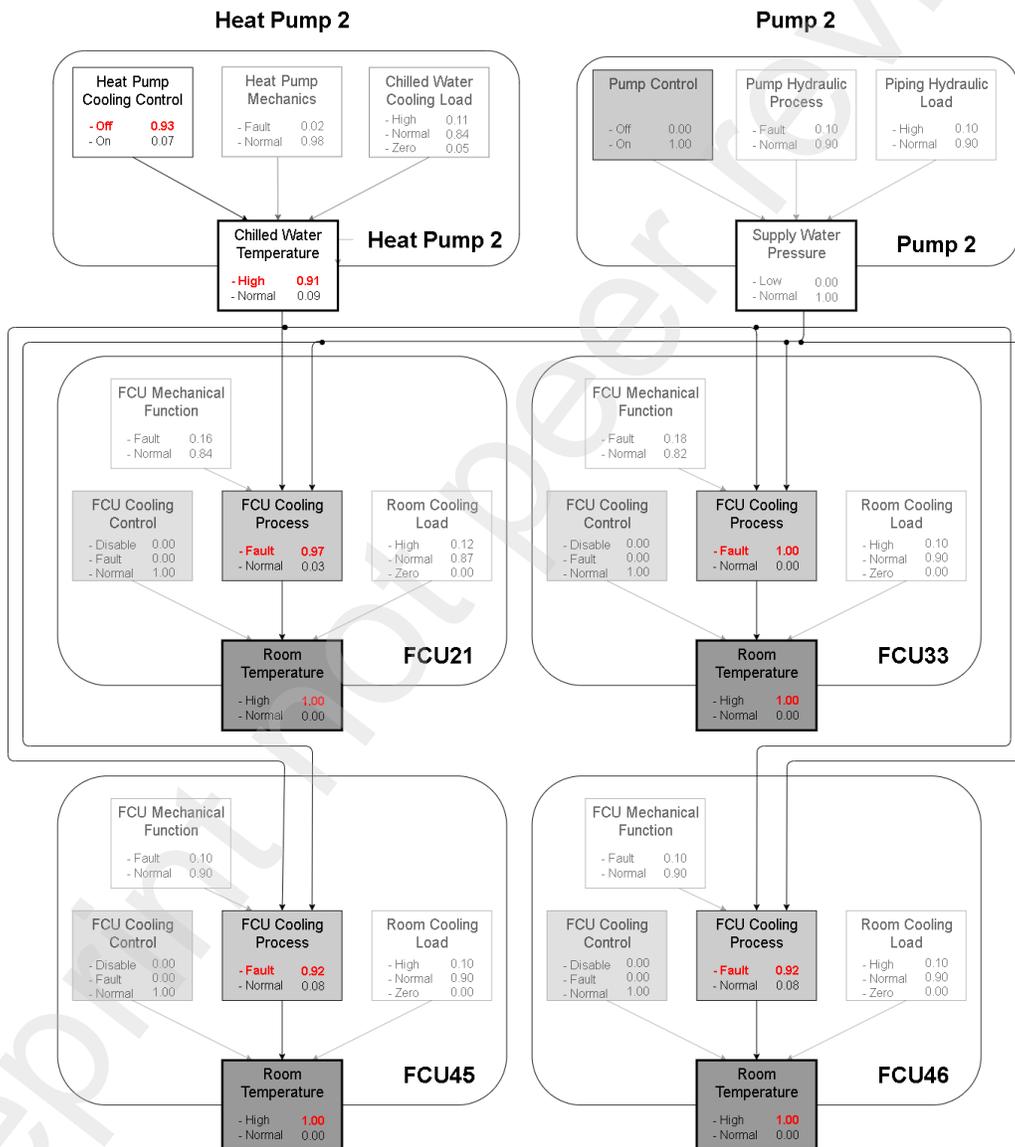
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Graphical Abstract

Integrated building fault detection and diagnosis using data modeling and Bayesian networks

Tianyun Gao, Sylvain Marié, Patrick Beguery, Simon Thebault, Stéphane Lecoecuche



Highlights

Integrated building fault detection and diagnosis using data modeling and Bayesian networks

Tianyun Gao, Sylvain Marié, Patrick Beguery, Simon Thebault, Stéphane Lecoecuche

- Transfer building system topology and expert knowledge to a Bayesian network.
- Transform equipment-level fault detection into building-wide fault diagnosis.
- Experiments on simulated and real-world buildings.

Integrated building fault detection and diagnosis using data modeling and Bayesian networks

Tianyun Gao^b, Sylvain Marié^a, Patrick Beguery^a, Simon Thebault^b,
Stéphane Lecoeuche^c

^a*Schneider Electric Industries, 160, avenue des Martyrs, Grenoble, 38000, France*

^b*CSTB, 24, rue Joseph Fourier, Saint-Martin-d'Hères, 38400, France*

^c*IMT Nord Europe, Institut Mines-Telecom, Univ. Lille, Centre for Digital Systems, 20
Rue Guglielmo Marconi, Villeneuve-d'Ascq, F-59650, France*

Abstract

Heating, ventilation, and air-conditioning (HVAC) equipment faults and operational errors result in comfort issues and waste of energy in buildings. An Automatic Fault Detection and Diagnosis (AFDD) tool could help facility managers fix comfort and energy issues more efficiently, by identifying the most probable root causes. Existing AFDD methods mostly focus on equipment-level fault detection and diagnostics ; almost no attention is given to building level fault diagnosis, considering inter-dependency between equipment through the energy distribution chain. In this work we propose a methodology to automatically derive a Bayesian network from HVAC system topology description such as Haystack. This Bayesian network models and estimates the state of all elements in the system, helping users to identify the most probable root fault. As it is able to ingest evidence from any source (field data, operators, or other models) and is capable of updating its estimates when new evidence is delivered, such a tool could have a great potential to be used interactively on the field. We applied the proposed methodology on simulated and real-world buildings and present in this paper one specific case.

Keywords: Bayesian network, Fault detection and diagnostics, Building, Descriptive metadata, Topology, Haystack, Data-driven modeling, BMS, HVAC

1. Introduction

Buildings, industry, and transportation are the three main energy consumers in our society. About 40% of energy is consumed in buildings in the European Union ([European Commission, 2020](#)). And within buildings, Heating, Ventilation and Air-conditioning (HVAC) systems consume the most energy.

HVAC system faults result in comfort issues and waste of energy. In order to help facility managers identify and fix faults more efficiently, it is essential to have an Automatic Fault Detection and Diagnosis (AFDD) tool, which is able to not only detect issues but also identify the root faults.

Today's Building Management Systems (BMS) are able to log data from all the sensors, actuators, and controllers of equipment in a building, which promises big potential in fault detection and diagnosis. In real-world HVAC and BMS systems, most alarms are derived from rule-based fault detectors embedded in HVAC Device controllers, and collected by the BMS. Existing AFDD methods mostly focus on equipment-level fault detection and diagnostics. See [Shi and O'Brien \(2019\)](#) for a complete review.

As a big picture of data-driven approaches, 3 main categories have been explored, at the sensor level with control charts able to detect and diagnose specific equipment faults using direct measurement ; at the signal level using statistics and processing methods to extract time series features able to characterize and discriminate varieties of faults ; at the level of the building dynamics, using multivariate model and analysis able to characterize more complex faults related to building dynamic changes. The first two categories have been extensively studied and [Gunay et al. \(2017\)](#) give a nice overview of almost 30 methods applied to detect faults and a wide list of AHU and VAV faults studied in the literature. The third category is more recently explored. As an example, [Turner et al. \(2017\)](#) develop a building HVAC fault detection method based on a data-driven approach and dynamic system modelling. This approach consists in a system identification technique based on recursive least-squares estimation of time-series data, able to model system dynamics and to detect residues between estimates and real measurements. [Ajib et al. \(2017\)](#) also develop a methodology based on data-driven models. Hybrid system dynamics are here estimated using a PWARX identification technique, able to extract normal and abnormal modes.

However, systems usually consist of a large number of devices and dynamic modes. In practice, analyzing the resulting large number of alarms

in order to trace root causes involves a significant amount of manual work and expert knowledge. This includes observing data trend logs (time series) of specific equipment at specific time periods, as well as navigating between associated equipment according to the system topology. Solving this problem automatically requires integrated analysis of the whole system, taking into account the inter-dependency between devices. In previous research, almost no attention has been given to such building-level fault diagnosis, considering inter-dependencies between HVAC devices through the energy distribution chain. We only found three interesting studies working in this area. [Schein and Bushby \(2006\)](#) proposed a hierarchical rule-based method to prioritize duplicated or conflicting alarms. [Verbert et al. \(2017\)](#) proposed a method based on Bayesian network to deal with equipment inter-dependency in HVAC systems. But this method highly depends on accurate and extensive building simulation to fit the conditional probability tables ; therefore is not very easy to deploy in practice. More recently, [Chen et al. \(2022\)](#) developed a new HVAC system cross-level fault diagnosis using a novel discrete Bayesian network (DisBN). This consists of causal relations among various components and sub-systems. It well addresses the challenge of cross-level fault diagnostics within a complex HVAC system. In this work, the network model is developed based on physical analysis and domain expert knowledge, but with a lack of automation exploiting existing topology and building meta-data.

Our study proposes a new building HVAC AFDD method using a Bayesian network to achieve building-level integrated fault diagnosis using operation data collected by BMS. Our methodology is based on the following proposal:

- A new systematic way of transferring building system topology information and expert knowledge to a Bayesian network able to infer the most probable root fault in the whole system based on comfort violation symptoms and operations data.
- A data-driven approach for extracting Bayesian parameters from BMS database and for designing a complete building-level fault diagnosis Bayesian network.

This paper presents this methodology and details its implementation. First a brief introduction of the methodology based on Bayesian network theory is given in section 2. Then the method for constructing the fault

diagnosis network from the building topology and from the BMS database is explained in section 3 with an example of its implementation on an HVAC system. Experiments of our methodology on simulated and real buildings are finally presented in section 4.

2. Motivation and methodology

2.1. Motivation

Building HVAC systems consist of heterogeneous equipment and systems, from room devices to central equipment, from heating and cooling emitters to hydraulic and air distribution systems. This represents an important technical database of various components, a variety of operating conditions (normal or abnormal, including faulty modes related to all possible degradation or faults) and a vast number of time series of various kind: setpoints, measurements, states, etc.

In this study, we propose to use a Bayesian network structure to represent the HVAC topology of a whole building and to achieve the diagnostic of possible faults using knowledge modeled by nodes and connections. Each node indicates the state (such as `Normal` or `Fault`) of an HVAC sub-system. The state of some nodes can be directly extracted from BMS data or provided by an operator from manual inspection, or also estimated with equipment-level data-driven fault detection algorithms (Gao, Tianyun et al., 2019). Updating the Bayesian network through inference leads to finding posterior distribution of unknown states, leading to finding root causes.

The choice of a diagnostic tool based on Bayesian network is based on the main following reasons:

- Bayes rule is a natural way of describing causal or probabilistic relationship between faults and symptoms. The parameters (conditional and prior probability) have clear statistical meaning, and the capability to output fault probabilities is more informative compared to binary fault detection results (normal or fault).
- Expert knowledge and building physics can be embedded into the definition of conditional and prior probabilities, so that the posterior probabilities calculation can reproduce a building manager reasoning more quickly and more efficiently.

- The graphical structure of a Bayesian network is able to mimic the building HVAC system energy chain and is flexible enough to cover most HVAC system topologies.
- Bayesian networks are good in dealing with uncertain, incomplete and even conflicting information, which is very common in building systems.
- Evidence about any node(s) can be provided as input to improve the current diagnosis, which makes it convenient for operators on site.
- Many efficient methods and tools exist to solve Bayesian network inference and belief updating problems.

2.2. Methodology

A Bayesian network structure representing a whole building system would contain thousands of nodes, which makes the task of creating or learning it challenging. Existing literature often limits the model scope to sub-systems with a few nodes, such as in [Verbert et al. \(2017\)](#) where a boiler-AHU-radiator system is modeled. We therefore propose a step-wise methodology, leveraging HVAC automation systems knowledge from standard building metadata, such as Project Haystack ([Haystack, 2014](#)).

Our methodology starts with defining a new “control flow diagram” descriptive model (section 3.1), that is both modular and extensible : building blocks can be simply assembled together to represent a complete system (section 3.2). We use this descriptive model as a pivot language to extract and structure the relevant HVAC system topology description (section 3.3) and to build the corresponding diagnosis Bayesian network (section 3.4), in a way that can be automated.



Figure 1: Methodology overview

We finally illustrate in section 3.5 the capability of the resulting network to diagnose root faults at system level based on given symptoms.

3. Design of an HVAC fault diagnosis network

This section is dedicated to the presentation of our methodology in the objective to design a building diagnosis tool.

3.1. Control flow model of a HVAC sub-system

An *HVAC sub-system*, such as a room radiator system, a hot water heating system or a ventilation system, is usually designed and operated to maintain a *controlled variable*, such as room temperature, hot water supply temperature, or air flow rate. The HVAC sub-system is regulated according to an *enable* (boolean) signal and a *setpoint* signal. Any generic HVAC sub-system can be represented by a control flow diagram as shown in Figure 2. It is composed of three components: the *Controller*, the *HVAC Device*, and the *Recipient System*.

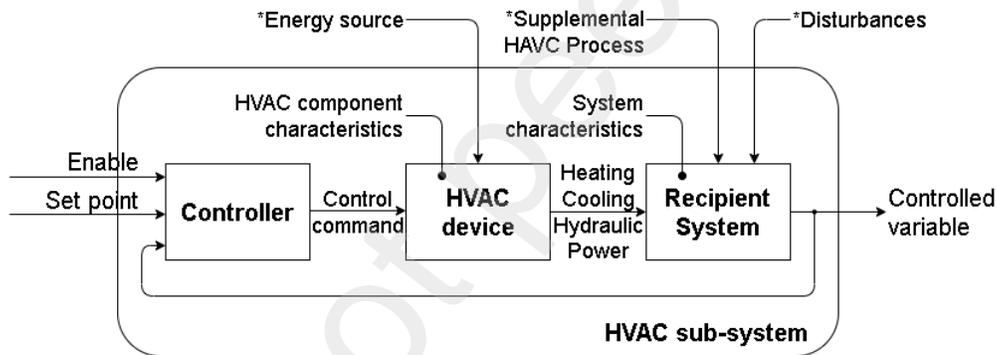


Figure 2: Generic control flow model of HVAC sub-systems

3.1.1. Controller

The *Controller* is the component regulating the *HVAC Device* to maintain the controlled variable at the *setpoint* (regulating controller), or simply switching on and off the *HVAC Device* (on-off controller).

Figure 2 shows a regulating close-loop controller. Controllers with other types of control loops (open-loop), output signals (two-position, floating, modulating), or control algorithms (dead-band, PI, PID, pulse width modulating) can be represented in a similar way ; see also Montgomery and McDowall (2008). The *enable* and *setpoint* signals usually come from an HMI (Human Machine Interface) either at the room thermostat or in the building

management system. They represent the control signals of the HVAC sub-system, and provide the basic criteria to evaluate whether the *Controlled Variable* is normal or abnormal.

3.1.2. HVAC Device

The *HVAC Device* refers to a heating / cooling / ventilating equipment. It is switched on/off or regulated by the *Controller* to give heating/cooling power or ventilation to the *Recipient System* to maintain the controlled variable (temperature, pressure, humidity, CO₂, etc.). The correct functioning of the *HVAC Device* depends on the *HVAC Component Characteristics* and the *Energy Sources* (if available).

3.1.3. Recipient System

The *Recipient System* refers to a room, a ventilation recipient system (air duct), a hot water recipient system (water tank and piping), etc. To maintain the controlled variable, the *Recipient System* needs power input from the *HVAC Device*. The amount of required power is related to the *Disturbances*, the *Recipient System* dynamics and the chosen setpoint. If the *Recipient System* is equipped with *Supplemental HVAC*, such as ventilation in addition to a room radiator system, the state of the *Supplemental HVAC* may also impact the *Recipient System*.

3.2. Connecting sub-systems to model a complete HVAC system

The control flow models of various HVAC sub-systems are presented in the [Appendix A](#). This list can be easily extended to develop a complete library covering most common types of HVAC Device.

By connecting corresponding component variables, we can easily build the control flow model for a whole HVAC system. We use a boiler-pump-radiator system as an example. The system is composed of one boiler, one hot water pump and three radiators serving three rooms individually. The boiler and the pump have on-off control, and the radiator has close loop regulating control. The control flow model of the full system is shown in [Figure 3](#).

In the next section we describe a methodology to construct such an entire model from the building topology descriptive metadata.

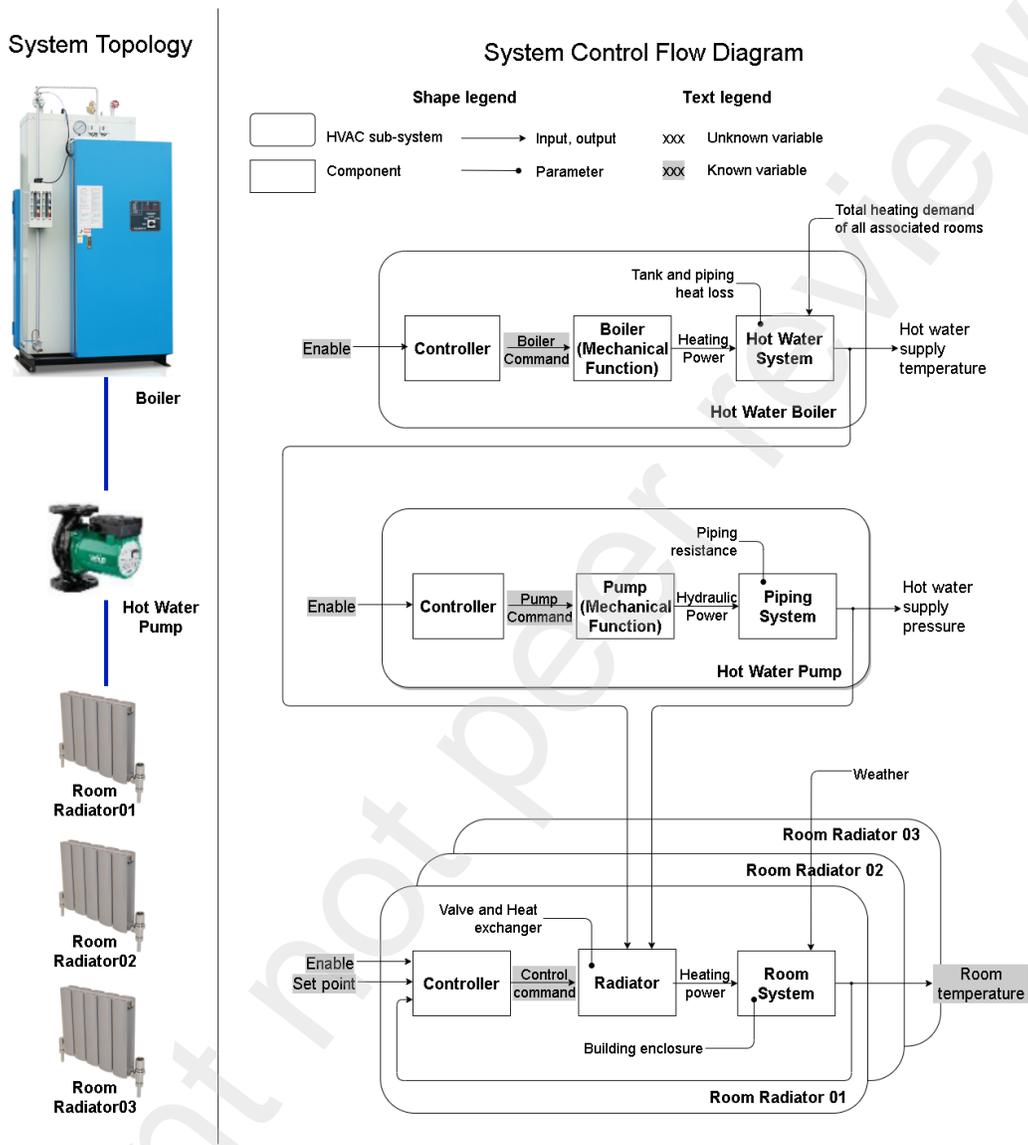


Figure 3: Boiler-pump-radiator system : Topology (left), Control Flow model (right)

3.3. Transforming building metadata into control flow model

In real world buildings, there are very often hundreds of equipment, and thousands of trend data logged in the BMS. It is a significant challenge to find the relevant data required to perform fault diagnosis of the complete system.

Over recent years, metadata models, taxonomies and ontologies such as Project Haystack (Haystack, 2014), IFC/BIM (ISO 16739-1:2018), Brick (Balaji et al., 2016) have been developed to standardize the description of building topology, HVAC systems, and associated trend data.

We have developed our method based on Haystack. With Haystack, a building is described by a list of entities that represent all of its constitutive elements: site, floor, room, equipment, measurement point, setpoint... Each entity is associated with an id (e.g. '@a-001'), descriptive tags that may contain a value (e.g. 'equip' or 'area: 55000ft²'), and relationships to other entities (e.g. 'chilledWaterRef: @a-07b8'). The complete building description can be exported to various formats such as CSV, or semantic web formats as JSON-LD or Turtle. An example of such description is illustrated in Figure 4 below. This description corresponds to the boiler-pump-radiator example presented previously in Figure 3. For readability purposes, relationships are indicated with arrows.

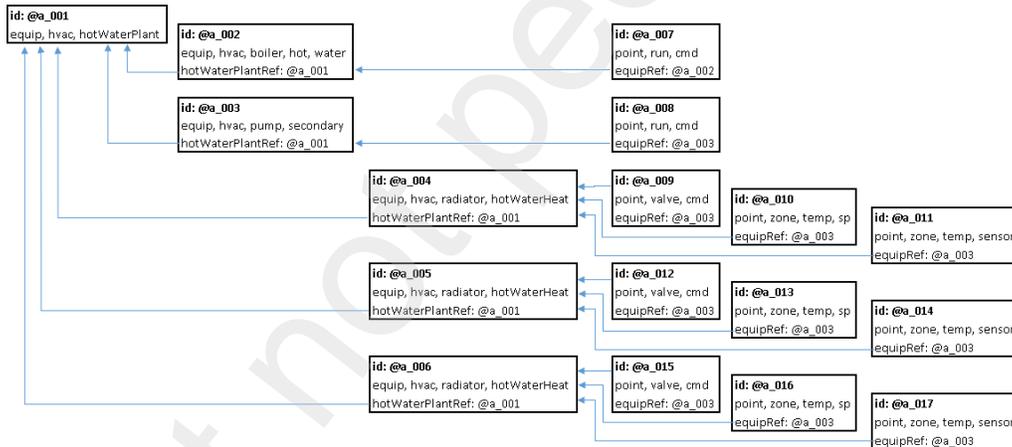


Figure 4: Haystack description of the boiler-pump-radiator example

We propose to leverage the Haystack description of a building to automate the process of creating the control flow model presented in 3.2. If the BMS data points are properly tagged using Haystack, these tags are automatically used to create the control flow diagram of the whole building. They are also used to associate the BMS data points to the corresponding signals in this diagram.

In the following we are going to use the boiler-pump-radiator example to illustrate how this process is done. [Figure 4](#) shows all the Haystack objects of this system. Converting this description into a control flow model consists in the following 3 steps:

Step 1: List HVAC sub-systems and find their type. From the Haystack metadata, get a list of all objects which have tag ‘hvac’ and ‘equip’, which means HVAC Device. Based on the additional tags, use [Table 1](#) to map each equipment to an HVAC sub-system model from the control flow library (see [Appendix A](#)).

Haystack tags	Sub-system in control flow model
radiator, hotWaterHeating	Room Radiator (Appendix A.1)
pump, secondary	Hot Water Pump (Appendix A.3)
hot, water, boiler	Hot Water Boiler (Appendix A.2)

Table 1: Creating HVAC sub-systems from Haystack

Step 2: Find all available signals for each HVAC sub-system. For each equipment entity ‘XY’, get a list of all entities which have tag ‘point’, and ‘equip=XY’. They represent the data points (timeseries) available for this equipment. Use the mapping table of the appropriate HVAC sub-system model (for example [2](#) for Room Radiator, [3](#) for Pump and Boiler) to map these entities to the control flow signal according to their available tags.

Haystack tags	Signal in control flow model
point, zone, air, effective, temp, sp	Setpoint
point, valve, cmd	Control command
point, zone, air, temp, sensor	Room temperature

Table 2: Signals in room radiator sub-system ([Appendix A.1](#))

Haystack tags	Signal in control flow model
point, run, cmd	Control command

Table 3: Signals in pump ([Appendix A.3](#)) and boiler ([Appendix A.2](#)) sub-systems

Step 3: Connect HVAC sub-systems. Based on the reference tags of each equipment, define the inter-relationship between the HVAC sub-systems. [Figure 4](#) indicates that the hot water of the three radiators ‘@a-004’, ‘@a-005’, and ‘@a-006’ are provided by the boiler ‘@a-002’ and pump ‘@a-003’.

With this methodology, we can now convert building descriptive data to control flow model, if the Haystack tags are correctly set.

3.4. From control flow model to fault diagnostics Bayesian network

In previous sections we described a control flow model ([3.1](#)) and a generic procedure to obtain it from a building description ([3.3](#)). We now describe how to build a Bayesian network according to the control flow model and perform fault diagnostics.

3.4.1. Bayesian network essentials

Bayesian networks are graph structures used for representing the probabilistic relationships among a large number of variables and for doing probabilistic inference (reasoning) with those variables. It was first developed by Judea Pearl in 1980s ([Pearl, 1988](#)). Nodes and edges form the *structure* of a Bayesian network. Each node represents a random variable. In this study we focus on *categorical random variables*. Directed edges are added from parent nodes to child nodes, to indicate direct influence. *Parameters* are given to each node to describe the prior (node without parent) or conditional (node with parents) probabilities. They are defined as prior probability tables (PPT) and conditional probability tables (CPT) respectively.

If the state of a node is observed, it is known as *evidence*. When we are absolutely sure about the state of the node, we set the probability of this state to one or zero. This is called *hard evidence*. Sometimes the information we have is not absolutely reliable, or the measurement we observe is subject to uncertainty. In this case we set the states to a probability between zero and one, and call it *uncertain evidence*. [Mrad et al. \(2015\)](#) provided a comprehensive review of uncertain evidence in Bayesian networks. There are two types of uncertain evidences ([Peng et al., 2010](#)):

- *Soft evidence* can be interpreted as “evidence of uncertainty”, and is represented as a probability distribution of one or more variables. This probability does not get updated in the belief updating process.

- *Virtual evidence* can be interpreted as “evidence with uncertainty”, and is represented as a likelihood ratio. It is used when one is uncertain about a claim of a hard evidence. In belief updating the posterior probability is calculated taken into account the likelihood ratio of this node and the states of other nodes in the network.

The process of computing the posterior distribution of variables given evidence is known as *probabilistic inference* (also formerly known as *belief updating*). Many inference algorithms exist, including exact methods such as variable elimination, poly-tree message passing and junction trees, and approximate methods such as Monte Carlo simulation and belief propagation. See [Neapolitan et al. \(2004\)](#); [Barber \(2012\)](#).

We now present how each aspect of the Bayesian network is created from the control flow model: Structure first (nodes and edges), then Parameters: prior probabilities and conditional probabilities.

3.4.2. Structure

Leveraging the generic control flow model of an HVAC sub-system illustrated in [Figure 2](#) we can easily build a sub-part of the fault diagnostics Bayesian network, abusively defined as “fault diagnostic Bayesian sub-network” and shown in [Figure 5](#). We now describe the method element by element, as for the generic control flow model (section [3.1](#)).

1. The *Controlled Variable* is represented by a node in the Bayesian sub-network named *Symptom* node, with state either `Normal` or `Abnormal`. The `Abnormal` state encompasses all situations where the controlled variable does not behave as expected by the system. Note that for convenience the `Abnormal` state may be renamed with a name related to the actual variable name. For example `Low` if the controlled variable is a Pressure, `Low` (resp. `High`) if it is a temperature during heating (resp. cooling) season, etc.
2. Each of the three blocks in the control flow model: *Controller*, *HVAC Device*, and *Recipient system*, also becomes one node in the Bayesian sub-network, named *Fault* node and representing its status:
 - the *Controller* node has states: `Normal` for normal behaviour; `Fault` for abnormal one; `Disabled` to take into account periods of time where the control is deactivated (typically based on the BMS Schedule),

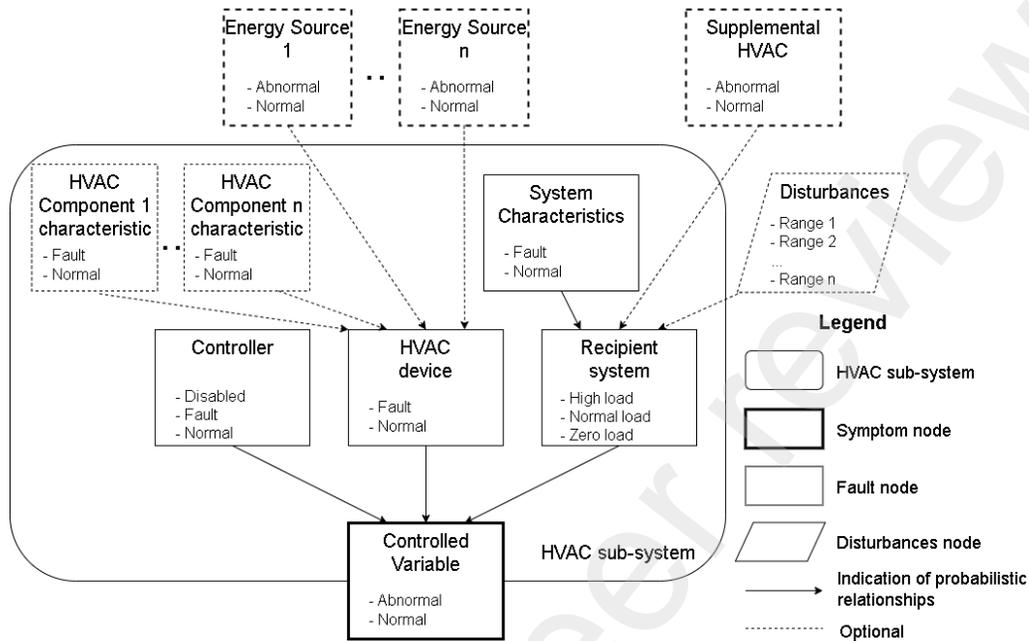


Figure 5: Fault detection Bayesian sub-network of an HVAC sub-system. Dotted lines indicate optional nodes.

- the *HVAC Device* node has states: **Normal** for normal behaviour; **Fault** for abnormal one,
- the *Recipient system* node has states: **Zero** for no load (for example when a room does not need heating because the external temperature and internal gains are sufficient); **Normal** for normal load (normal need for heating or cooling); **High load** for abnormally high loads (for example when a window is left open in a room and the heating system cannot compensate).

3. The vertical arrows on top of the blocks in the control flow model represent properties and external inputs. Each of these signals becomes an independent parent node of its associated block node in the Bayesian sub-network:

- For the *HVAC Device* node (e.g. Radiator), **Fault** state is likely to be caused by abnormal *HVAC Component characteristic* (e.g. Radiator valve, heat exchanger) or abnormal *Energy source* (e.g. Supply Water Temperature, Supply Water Pressure). The first is represented by a

Fault node with states **Normal** and **Fault**. The latter is represented by the *Controlled Variable* node of another Bayesian sub-network.

- For the *Recipient system* node (e.g. Room), the **High/Normal/Zero** load state is a function of the *System Characteristics* (e.g. the Building Enclosure), *Disturbances* (e.g. Outdoor air temperature), and *Supplemental HVAC* (if available) states. The first two are represented either by a *Fault* node with states **Normal** and **Fault**, or by a *Disturbance* node with ad-hoc states. The latter is represented by the *Controlled Variable* node of another Bayesian sub-network.

3.4.3. Structure properties

The choices described in 3.4.2 lead to the following desirable properties:

- **Minimal need for parameter learning.** The structure of the Bayesian network is purposely designed to minimize the necessity of estimating (or learning) and tuning the prior and conditional probabilities. Indeed, most of the state variable dependencies are binary logic, which means the conditional probabilities are either 1 or 0. For example when a controller has **Fault** state, controlled variable is **Abnormal** with a probability of 1. The only conditional probability that needs to be estimated based on expert knowledge is the one of the *Recipient System* node. The associated probability tables (PPTs and CPTs) and other parameters are detailed in 3.4.4 and in 3.4.5.
- **Composability.** The above standard design of Bayesian sub-network structure is used to model each HVAC sub-system in the global system. In the objective to design a diagnosis tool based on a complete Bayesian network, our method consists in connecting Bayesian sub-networks, mirroring connections between HVAC sub-systems that compose the HVAC system in a building. The connections of different sub-systems are built between the *Controlled Variable* node of the energy sources and the *HVAC Device* node of the energy consumers. As an example, the complete fault diagnosis network of the boiler-pump-radiator topology is given in Figure 6.

Finally, note that the horizontal arrows in the control flow model represent input and output signals for the three blocks *Controller*, *HVAC Device*, and *Recipient system*. These do not directly appear in the Bayesian sub-network

structure, but we will see in next section that the conditional probability tables (CPT) are designed to reflect these connections. In addition, the behavior of the outputs reacting to the inputs can be used by an expert or an algorithm to reveal the state of the blocks (generate evidence).

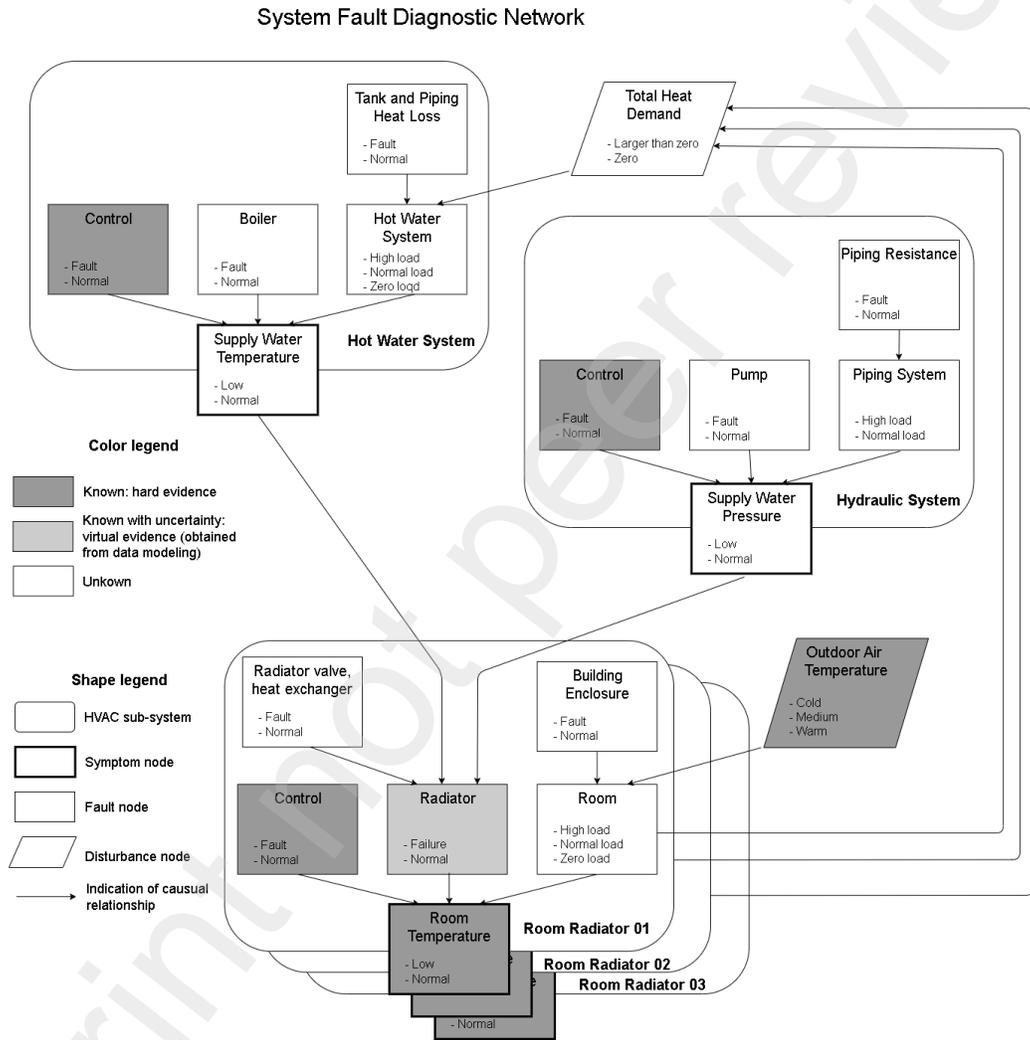


Figure 6: Boiler-pump-radiator system – fault diagnosis Bayesian network

3.4.4. Parameters: prior probabilities

In the HVAC diagnosis sub-network as shown in Figure 5, or network as shown in Figure 6, the nodes without parents are root cause nodes. We

propose below two strategies to set prior probabilities for these nodes, depending on their kind: *Disturbance* nodes or *Fault* nodes.

1. *Disturbance* nodes refer to variables such as outdoor air temperature, total heating demand of the hot water system, etc. They reflect the usage context of the system. Their prior probability distribution can for example be obtained from historical data, or expert knowledge. Table 4 illustrates a prior probability table for a *Outdoor air temperature* node corresponding to a classical yearly distribution in Europe. In real-life systems, this prior probability table (PPT) should be updated periodically to better fit to actual conditions (e.g. one distinct probability table for heating/mid-season/cooling seasons or quarterly, monthly, weekly...). Finally note that some *Disturbance* state variables can be actually monitored. The state derived from live observations can be posted as evidence on the node to further refine its status.

	Lower than 10°C	Between 10°C and 20°C	Larger than 20°C
Outdoor air temperature	0.4	0.2	0.4

Table 4: Prior probabilities of the disturbance node in the boiler-pump-radiator example

2. *Fault* nodes correspond to the failure of a specific equipment or building component. The relative importance of prior probabilities set for the various *Fault* nodes across the entire network has a significant impact on the final diagnosis result (most probable fault for a given set of symptoms); these should therefore be carefully defined. The prior probability of equipment fault can be derived from a failure rate model if one is available, such as the well-known bathtub failure rate (see Signoret and Leroy (2021) or Finkelstein (2008) for a review).

When no such model is known, prior probabilities of faults may be estimated based on field experience. In the boiler-pump-radiator example, the prior probability of all root fault nodes are empirically set to 0.1.

3.4.5. Parameters: conditional probabilities

All nodes with parents have to be associated with conditional probability tables (CPT) representing their probabilistic relationship with the parent nodes. We propose below one strategy for each type of node, to define these tables.

1. *Symptom* nodes represent the fault status of controlled variables. We propose to assume a logical causal relationship between the states of these nodes and their parents, encoded by the following expert rules:

- When the load of the *Recipient System* is **High** (higher than the maximum power output of the *HVAC Device*), the *Symptom* node is **Abnormal** (e.g. temperature is **Low** in heating season) in any case, as the controlled variable can't possibly reach the normal range even if the entire HVAC system is working properly.
- When the load of the *Recipient System* is **Normal** (above zero and smaller than the maximum power output of the *HVAC Device*), the controlled variable needs to be regulated by a working HVAC system to be normal. Therefore the *Symptom* node is only **Normal** when the *Controller* and the *HVAC Device* are both **Normal**. A *Controller* fault (including disability) or a *HVAC Device* failure will all cause *Symptom* node to be **Abnormal**.
- When the load of the *Recipient System* is **Zero**, the controlled variable does not need to be regulated to be normal. The *Symptom* node is therefore **Normal** in any case. This is the case typically for room and air heating or cooling systems in transition seasons. This however does not apply to Air fans and pumps: they usually don't have 'zero load' status, since power is always needed to maintain water pressure and air pressure.

The above proposed logic translates into the conditional probability table in [Table 5](#).

System	Controller	HVAC	Controlled Variable	
			Abnormal	Normal
High	Fault	Fault	1	0
High	Fault	Normal	1	0
High	Normal	Fault	1	0
High	Normal	Normal	1	0
Normal	Fault	Fault	1	0
Normal	Fault	Normal	1	0
Normal	Normal	Fault	1	0
Normal	Normal	Normal	0	1
Zero	Fault	Fault	0	1
Zero	Fault	Normal	0	1
Zero	Normal	Fault	0	1
Zero	Normal	Normal	0	1

Table 5: Conditional probabilities of *Symptom* nodes

2. For *HVAC Device* nodes we also propose to assume a logical causal relationship between the states of these nodes and their parents, encoded by the following expert rule: if *HVAC Component Characteristics* has state **Fault** or *Energy Source* (when applicable) has state **Abnormal**, then *HVAC Device* has state **Fault**. For example in the case of the room radiator system, *HVAC Component Characteristics* has state **Fault** when the valve is stuck close or the heat exchanger is blocked. *Abnormal Energy Source* refers to either abnormal hot water temperature or abnormal pressure, which are the *Symptom* nodes of two other HVAC sub-systems (Hot water sub-system and Hydraulic sub-system), and therefore are themselves caused by other faults. By connecting the *HVAC Device* node to the *Energy Source* parent node, the inter-dependency between HVAC sub-systems, or the causal relationship between faults in different HVAC sub-systems are represented in the Bayesian network.

The above proposed logic translates into the conditional probability table in [Table 6](#).

HVAC component characteristics	Energy source	HVAC Device	
		Abnormal	Normal
Fault	Abnormal	1	0
Fault	Normal	1	0
Normal	Abnormal	1	0
Normal	Normal	0	1

Table 6: Conditional probabilities of *HVAC Device* nodes

3. For *Recipient System* nodes we propose to learn their conditional probabilities from historical data. In correctly sized HVAC sub-systems, **High** recipient system load is caused by **Abnormal System Characteristics**, subject to certain conditions related to *Disturbances*. In the room radiator system example, *Disturbances* refer to weather (outside temperature, solar radiation, etc.) and internal heat gains, *System Characteristics* refer to the building enclosure heat transfer coefficient and heat capacity. If the radiator has the right size, in normal situations, the heat capacity should always be able to cover the system load. High system loads are caused by building enclosure fault, such as an open window or a damaged insulation, and only appears when the outside temperature is low.

The conditional probability table of *Recipient System* nodes can be learned from disturbances data and corresponding system load data in normal case and fault case. An alternative when historical data is missing is to simplify the network by removing the *Disturbance* and *System characteristics* nodes, such that only prior distribution needs to be learnt for *Recipient System* from past data.

Table 7 gives an example of the heat load distribution of building enclosure **Normal** (window closed) and **Fault** (window opened) case, for the boiler-pump-radiator example.

Building enclosure (<i>System Character.</i>)	Outdoor air temperature (<i>Disturbances</i>)	Room (<i>Recipient System</i>)		
		High load	Normal load	Zero load
Fault	Lower than 10°C	0.85	0.15	0
Fault	Between 10°C and 20°C	0.4	0.4	0.2
Fault	Higher than 20°C	0	0	1
Normal	Lower than 10°C	0	1	0
Normal	Between 10°C and 20°C	0	0.5	0.5
Normal	Higher than 20°C	0	0	1

Table 7: Conditional probabilities of *Room* nodes in the boiler-pump-radiator example.

3.5. Fault diagnosis decision

A big advantage of Bayesian network is its flexibility in data availability: any node can receive evidence, and all available evidence is leveraged for inference. Fault diagnosis can thus be performed in an incremental fashion: users or systems provide evidence, then inference is executed to get the probability of all root fault nodes and find out the most possible one. New evidence can then be posted and the probabilities updated, etc. We propose that indubitable observations such as low temperature are posted as hard evidence, while uncertain observations or results, such as individual equipment fault detection results from another system or algorithm (when available) are posted as *virtual evidence*.

We now present an example of such a diagnosis process, starting with an observed symptom: a low room temperature in room 1, posted as hard evidence on the associated *Symptom node* of the network (Case 1). We then consider situations with an increasing amount of evidence posted into the network, mimicking the progress of an assisted human inspection (Cases 2 to 5). Inference results after each new evidence are shown in Table 8, one column per inference run. Room 1 Temperature Symptom node is listed, as well as Symptom nodes of the sub-systems which are served by the same energy source (room temperature of other rooms). Updated probabilities are provided for all root fault nodes and intermediate fault nodes that are direct or indirect parents of the symptom nodes.

	Case1	Case2	Case3	Case4	Case5	Case5'
Symptom nodes						
<i>Room radiator 01</i>						
Room temperature	Low	Low	Low	Low	Low	Low
<i>Room radiator 02</i>						
Room temperature			Normal	Normal	Normal	Normal
<i>Room radiator 03</i>						
Room temperature			Normal	Normal	Normal	Normal
Disturbance nodes						
Outdoor air temperature		Medium	Medium	Medium	Medium	Medium
Virtual Evidence						
<i>Room radiator 01</i>						
Radiator fault					0.8	0.2
Intermediate fault nodes						
<i>Room radiator 01</i>						
Radiator fault	0.86	0.86	0.70	0.79	0.94	0.48
Room high load	0.14	0.13	0.19	0.27	0.13	0.56
<i>Hot water system</i>						
Hot water temperature low	0.44	0.44	0.29	0.26	0.30	0.16
Hot water system load high	0.15	0.15	0.10	0.13	0.15	0.08
<i>Hydraulic system</i>						
Hot water pressure low	0.45	0.45	0.30	0.27	0.32	0.16
Hydraulic system load high	0.17	0.17	0.11	0.14	0.17	0.09
Root fault nodes						
<i>Room radiator 01</i>						
Controller fault	0.17	0.17	0.25	0	0	0
Valve heat exchanger fault	0.17	0.17	0.25	0.36	0.43	0.22
Building enclosure fault	0.16	0.20	0.25	0.33	0.20	0.59
<i>Hot water system</i>						
Controller fault	0.17	0.17	0.11	0	0	0
Boiler mechanical fault	0.17	0.17	0.11	0.14	0.17	0.09
Tank piping heat loss fault	0.16	0.16	0.11	0.14	0.16	0.09
<i>Hydraulic system</i>						
Controller fault	0.17	0.17	0.11	0	0	0
Pump mechanical fault	0.17	0.17	0.11	0.14	0.17	0.09
Piping resistance fault	0.17	0.17	0.11	0.14	0.17	0.09

Table 8: Fault diagnosis inference results. Posted evidence is marked in gray, changes with previous case in bold font. Root faults are marked in red (most probable: dark red ; less probable: light red).

As we can see, with only the ‘Room 1 low temperature’ evidence, all possible root faults have similar probabilities (Case1). We can check the **outdoor air temperature** to see if it is the direct cause. Observing ‘Medium’ external temperature does not change the diagnosis as this is not a direct valid

reason for a low temperature in rooms (Case2). Once it is known that the **other rooms** served by the same hot water system all have normal temperature, although the hot water system temperature and water flow information are not available, the Bayesian network is able to infer that the root fault is most probably located in room 1 (Case3). By observing the room temperature **control signals**, we can identify whether they are normal or not. Assuming that the valve control signal is already at maximum, which is correct reaction to the temperature deviation to the setpoint, we can set the controller fault node probability to zero. Now the most probable fault is either radiator mechanical fault (valve or heat exchanger), or building enclosure fault (such as the window in room 1 is open) (Case4). By observing how room temperature is reacting to the heating valve opening degree, we can create virtual evidence representing how likely it is that the radiator has a **mechanical problem**: likely (0.8, Case5) or not (0.2, alternate Case5'); posting this evidence in the network results in the correct fault being isolated.

4. Experiment and discussion on real case studies

In this session we apply our methodology to build the fault diagnosis network of HVAC system of one real office building 'Retz' located in Nantes, France. The building is equipped with heat pumps and air handling units (AHU) as central HVAC devices, and with floor heating, fan coil units (FCU) and variable air volume box (VAV) in the rooms. The system topology is shown in Figure 7. There are 55 rooms, more than 800 points with data trends in total.

The corresponding building information Haystack model is illustrated in Figure 8. Because of the space limitation, the points are not included in the diagram.

4.1. Network structure and parameters

We create the fault diagnosis network based on the Haystack model, following the procedure described in 3.4. We removed the AHUs to simplify the Bayesian network structure for the purpose of the experiment, to focus on the FCUs in rooms, the associated heat pumps and the distribution pumps. We manually checked that removing AHUs had no impact in the timespan of the dataset. As heating and cooling seasons are independent we model each equipment by two HVAC sub-systems (two Bayesian sub-networks) representing heating and cooling functions respectively, and shar-

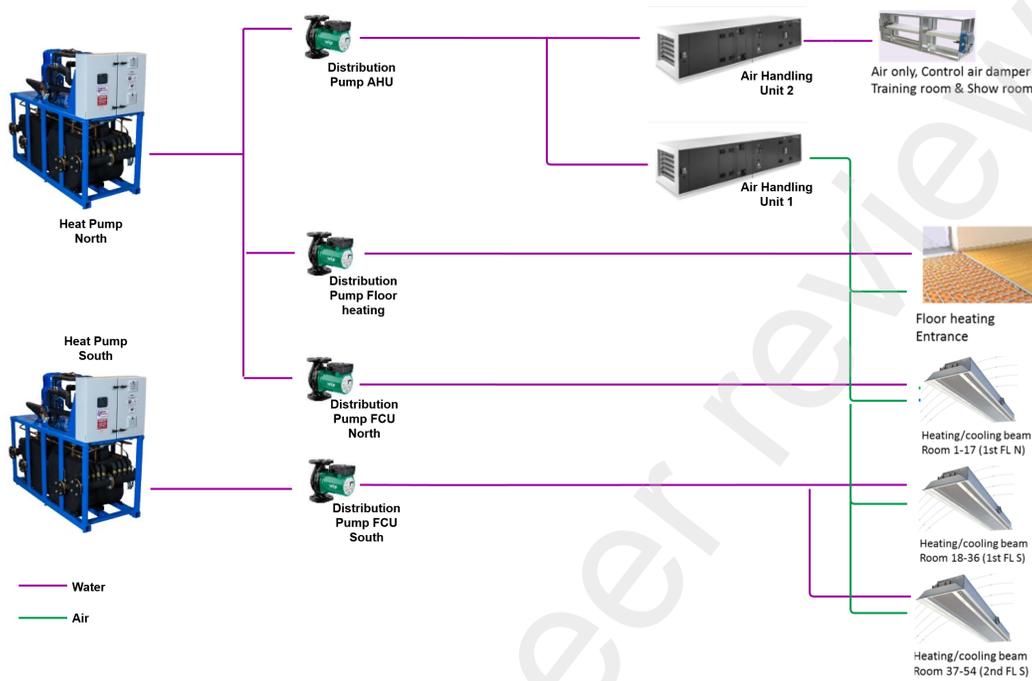


Figure 7: HVAC system topology of Retz building

ing common *Recipient System* and *HVAC Device Characteristic*. The fault diagnosis network is shown in Figure 9.

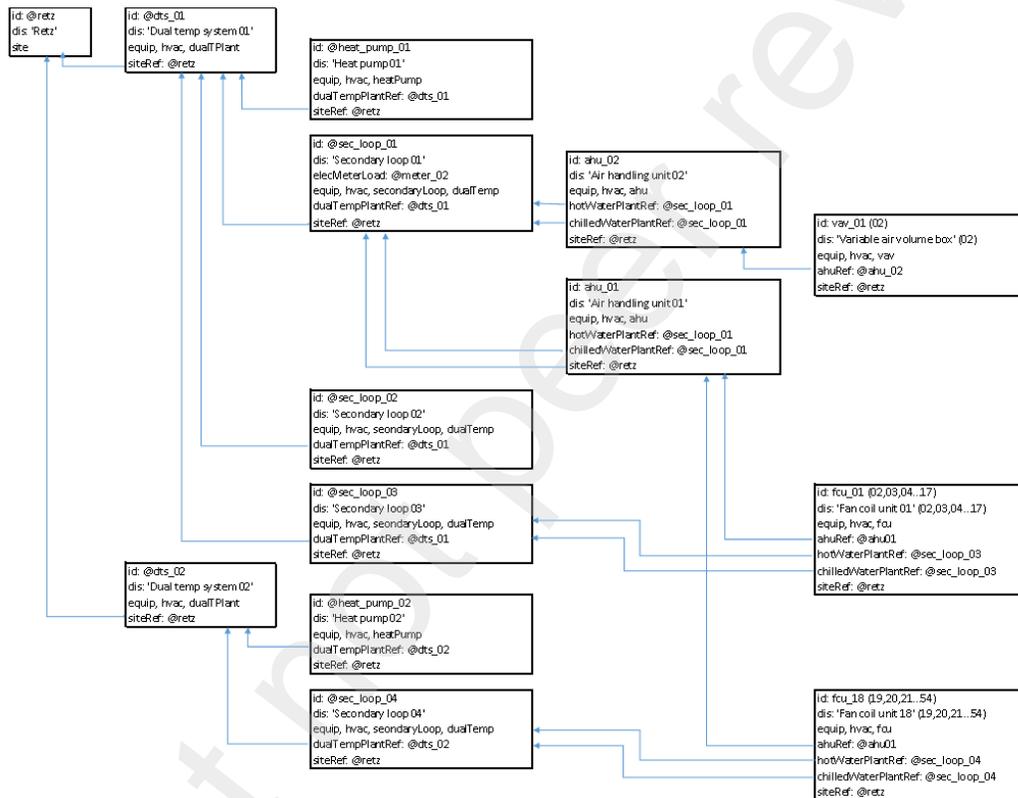


Figure 8: Haystack model of the Retz building

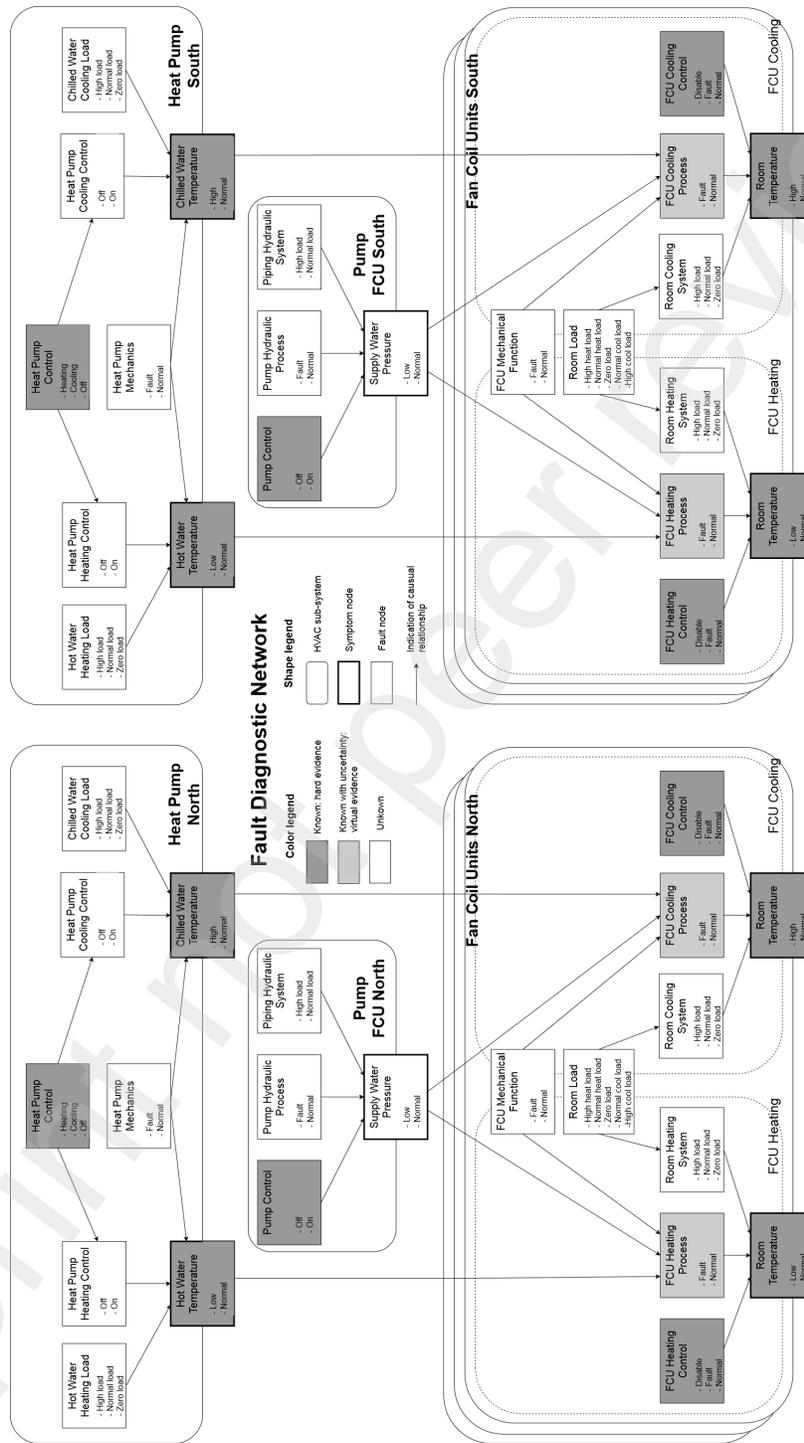


Figure 9: Fault diagnosis network of the 'Retz' building

4.2. Evidence and inference results

In this case study, hard evidence is obtained every hour from real operation supervision data. The state of controlled temperature variables is obtained by comparing the actual room temperature to the setpoint during building operation hours with a 0.5°C tolerance. The general heat pump controller state is obtained from the heating season status. Pump controller states are obtained from the corresponding data point. Finally, we simulate availability of virtual evidence about the Fan Coil Units with a pre-recorded evidence signal.

Probabilities of all possible fault nodes are calculated through inference to find out the most probable root fault cause. The whole process is executed automatically in a Python program using the SMILE reasoning engine from BayesFusion, through its PySMILE Python wrapper (Tungkasthan et al., 2010). Fault diagnostics for one specific hour takes less than a second. Going through a whole year of data takes about half an hour, which is reasonable speed.

Using this method, we have successfully revealed some issues, even with incomplete historical data. Below is one example.

On July the 6th, during the cooling season, heat pump data missing. It is unknown whether the heat pumps were running or not. We now investigate this case to see if the Bayesian network is tolerant to missing data and still able to correctly identify that heats pumps are not running correctly.

The symptoms are abnormally high temperatures in rooms 21, 33, 45, and 46. An extract of the inference results for these four rooms is shown in Figure 10. In all four rooms, the valve correctly opens to cool down and maintain room temperature at the setpoint. Based on this evidence as well as all others, the Bayesian network was able to correctly identify that the heat pump is not in cooling mode (probability 90%). This indicates that the high temperature in four rooms are caused by the conflict between room FCU and heat pump heating / cooling mode.

4.3. Other experimental results

We conducted two other experiments to test the generalization capabilities of our approach.

- with a "digital twin" simulation model of one Schneider Electric building (GreenOValley 38TEC-T11) in Grenoble. The model was created with IDA-ICE and simplified by removing some rooms and replacing

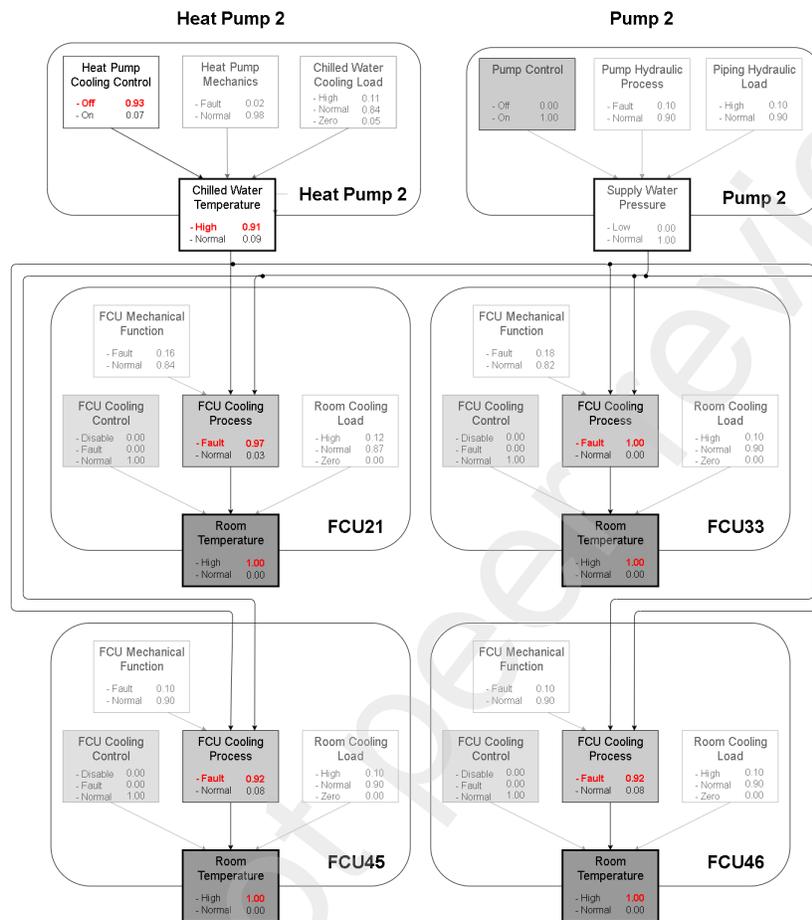


Figure 10: ‘Retz’ building, building level diagnosis inference results with missing data: heat pump off causing high room temperatures.

the heat pump models with simple boiler and chiller models. The final model results in 56 FCU-equipped rooms, 3 AHUs, 1 chiller with 2 pumps and 1 boiler with 2 pumps.

- with the real GreenOValley 38TEC-T11 building

In both cases we were able to successfully describe the entire system using the control flow approach (Figure 11) and to use resulting Bayesian networks to detect root faults based on available evidence (Figure 12).

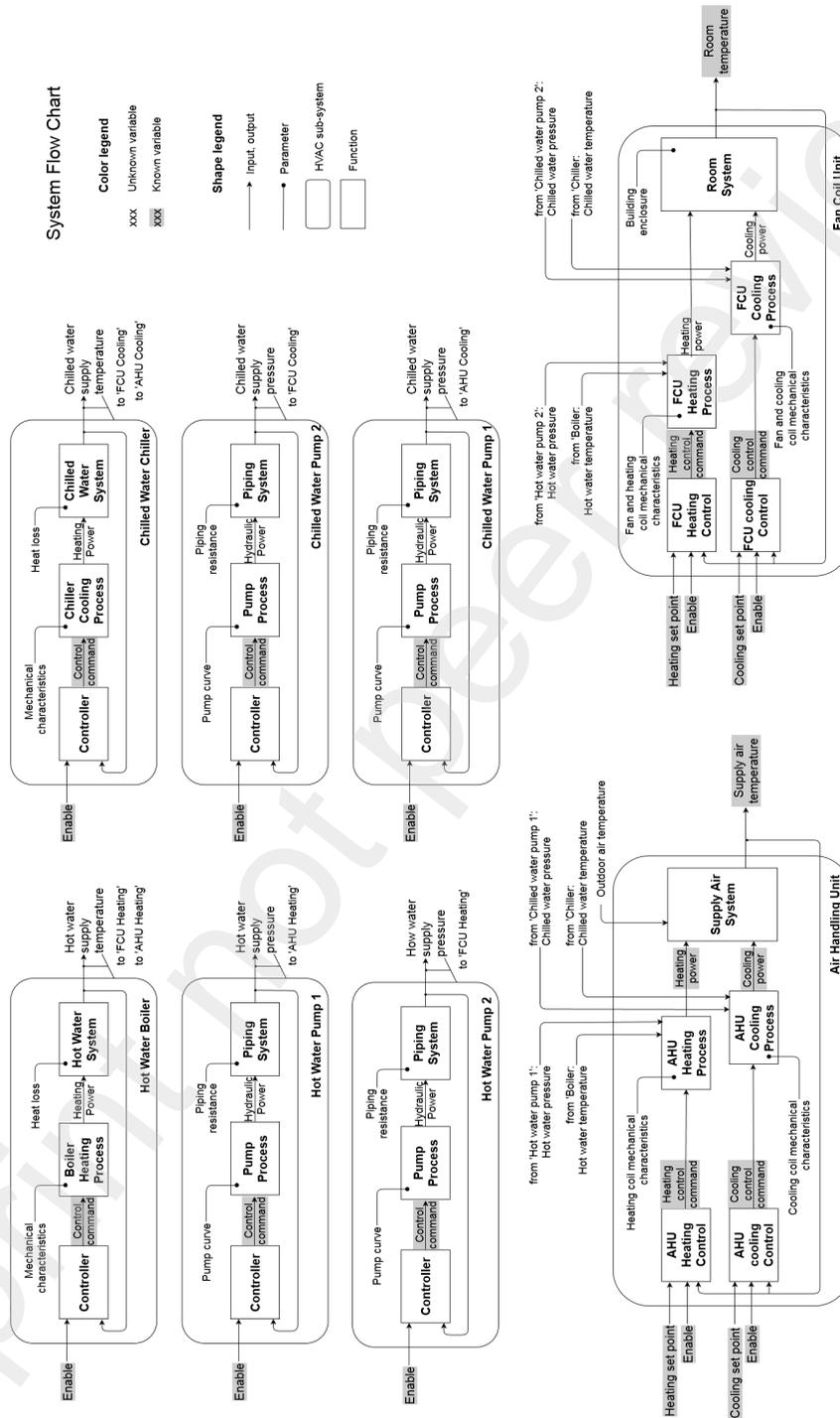


Figure 11: Simulated GreenOValley building, control flow diagram.

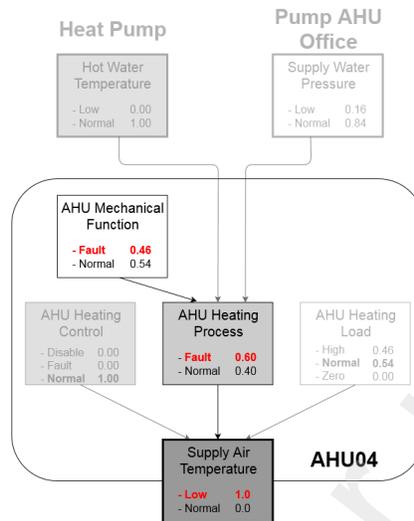


Figure 12: GreenOValley building, building level diagnosis inference example: AHU04 mechanical fault (valve) causing low supply air temperature

5. Conclusion

In this study we propose a new method to design a diagnosis tool for HVAC faults at the whole building level, based on the use of a Bayesian network. This Bayesian network has a flexible and modular structure, with “Bayesian sub-networks” modeling HVAC sub-systems so that whole systems can be represented easily. As this construction heavily relies on expert knowledge, we propose a generic “control flow” system description that makes it easier for experts to describe sub-systems, and that is used as an intermediate step in the Bayesian network construction. An initial library of such “control flow” sub-systems, used in this study, is provided in [Appendix A](#). Assuming that such a library is available and that the building topology is correctly described – for example by Haystack tags – we propose a step-wise approach to create the complete Bayesian network in an automated way. Its parameters (PPTs and CPTs) are defined either from rules or historical data. The resulting model fuses different type of data from sensors to field observations, and can integrate virtual evidence for example from other fault detection systems. It is able to deal with inter-dependencies between various HVAC devices in order to perform global fault diagnosis and isolation of

root cause faults. The method has been tested on data from real and simulated office buildings. It revealed to be flexible with HVAC system topology and data availability, requires small computational effort, and provides good diagnosis accuracy.

To enable this tool to be used operationally, remaining work need to be done, as :

- an exhaustive “control flow” library of possible HVAC Systems (see [Appendix A](#)), associated with a clear way to map them from the Haystack topology description (similar to part [3.3](#))
- a prior data quality check and pre-process, in particular on time series data to avoid diagnosis errors
- an "interactive wizard" tool for the Facility Manager, with intuitive user-friendly results display overlaid on the actual system synoptics, as well as easy-to-enter field evidence inputs

In real-world systems, manual inspection of room temperatures or machine states is not always simple and fast ; it can become intractable as the building size grows, or even unfeasible for state variables that are not monitored. Future work will explore how users could leverage data-driven methods – in particular equipment-level fault detection models – to derive virtual evidence and eventually ease the diagnosis process. For example, using supervised machine learning regression models to detect faults on Air Handling Units ([Gao, Tianyun et al., 2019](#)) and generate virtual evidence.

6. Acknowledgements

The authors would like to thank Henri Obara and the Facility Management team from GreenOValley 38TEC for their help and feedback.

Appendix A. Examples of specific HVAC sub-systems

In this section, the generic control flow model is created for several specific HVAC sub-system, and the fault diagnostics Bayesian sub-networks are derived accordingly, following the methodology presented in section 3. Finally, a summary is given in Table A.1.

Appendix A.1. Room Radiator

Figure A.1 illustrates the control flow model of a room radiator sub-system. The radiator valve is regulated by the controller according to the room temperature and the setpoint. The heating power provided by the radiator to the room depends on the valve status and heat exchanger, but also on the hot water supplied as input to the radiator. Finally, the amount of heat required by the room is dependent on weather and building enclosure. The associated fault diagnostic Bayesian sub-network is illustrated in Figure A.2.

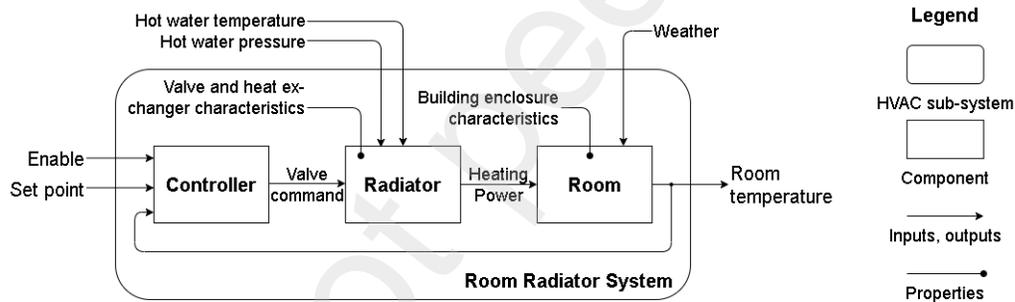


Figure A.1: Control flow model of a room radiator sub-system

Appendix A.2. Hot Water

Figure A.3 illustrates the control flow model of a hot water sub-system. The boiler is regulated by the controller to maintain the supply water temperature at the setpoint. The boiler gives heating power to the hot water system to cover the heating demand of all associated rooms and the tank and piping heat loss. The associated fault diagnostic Bayesian sub-network is illustrated in Figure A.4.

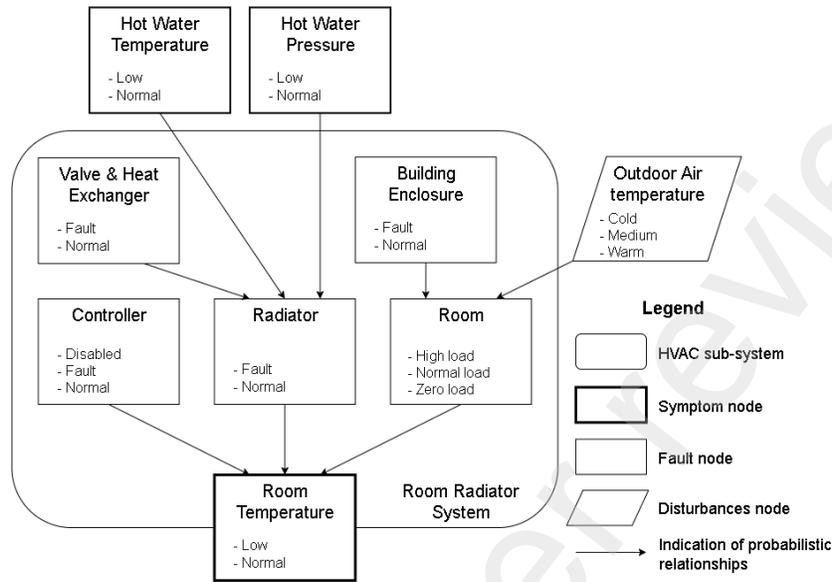


Figure A.2: Fault diagnosis Bayesian sub-network of a room radiator sub-system

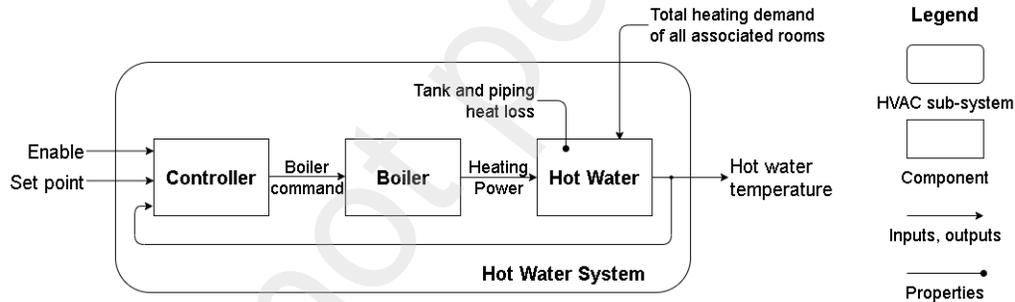


Figure A.3: Control flow model of a hot water sub-system

Appendix A.3. Hydraulic

Figure A.5 illustrates the control flow model of a hydraulic sub-system. The pump is regulated by the controller to maintain the supply water pressure at the setpoint. The pump gives hydraulic power, which refers to the combined pressure and flow rate the pump is able to provide to the piping, as can be illustrated in a pump curve shown in Figure A.7.

The pressure and corresponding flow rate is given by the pump speed. Disturbances are the resistances given by the position of all the valves in the piping. If the piping system is blocked or has leakage (*Piping Resis-*

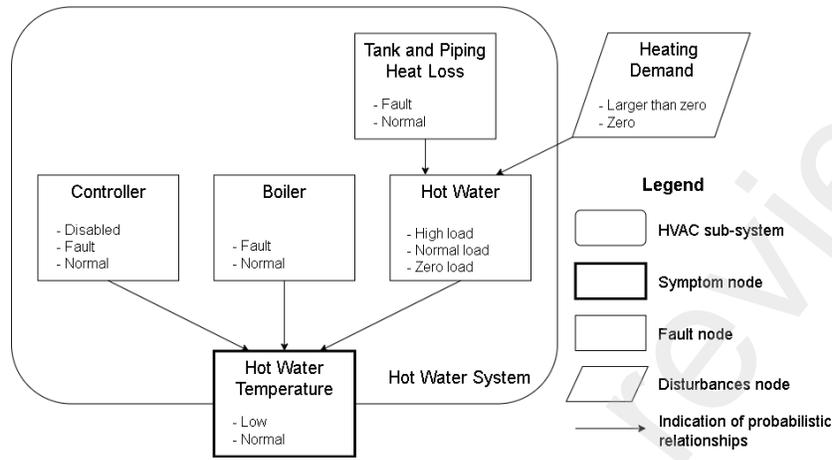


Figure A.4: Fault diagnosis Bayesian network of a hot water sub-system

tance status is Abnormal), the pump would not be able to provide enough power to maintain the water pressure and water flow (*Piping System* status is High load). The associated fault diagnostic Bayesian sub-network is illustrated in Figure A.6.

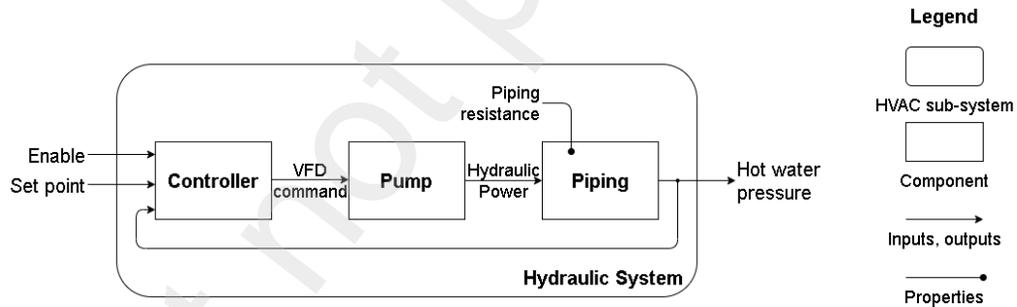


Figure A.5: Control flow model of a hydraulic sub-system

Appendix A.4. Ventilation

Figure A.8 illustrates the control flow model of a ventilation sub-system. The ventilation power refers to the combined pressure and flow rate the air fan is able to provide. It is defined by the fan curve (similar to the pump curve). If the duct is blocked or has leakage (*Duct Resistance* status is Abnormal), the fan would not be able to provide enough power to maintain the air

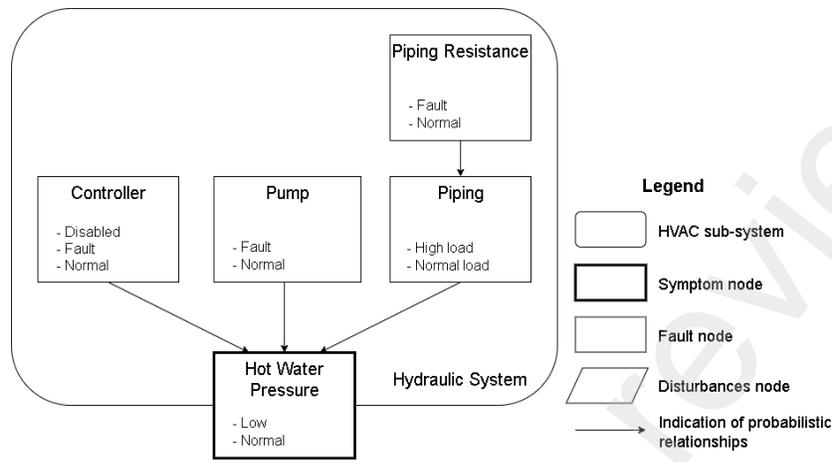


Figure A.6: Fault diagnosis Bayesian sub-network of a pump hydraulic sub-system

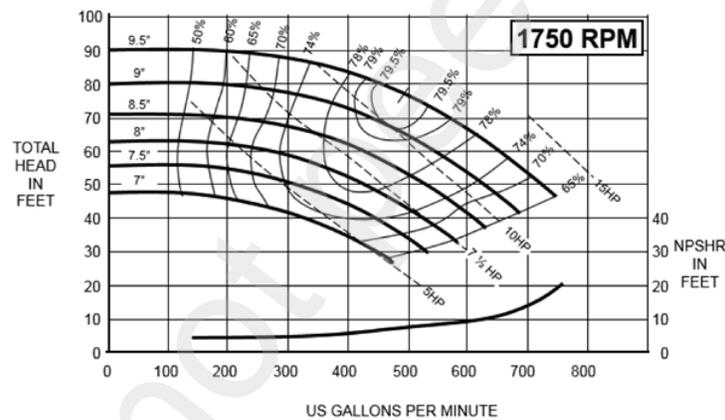


Figure A.7: Example of VFD pump curve

pressure and air flow (*Duct System* status is High load). The associated fault diagnostic Bayesian sub-network is illustrated in [Figure A.9](#).

Appendix A.5. Chilled Water

[Figure A.10](#) illustrates the control flow model of a chilled water sub-system. The chiller is regulated by the controller to maintain the chilled water temperature at the setpoint. The chiller gives cooling power to the chilled water system. The cooling function is dependent on the functioning of the compressor, heat exchanger, and expansion valve, as well as the condenser

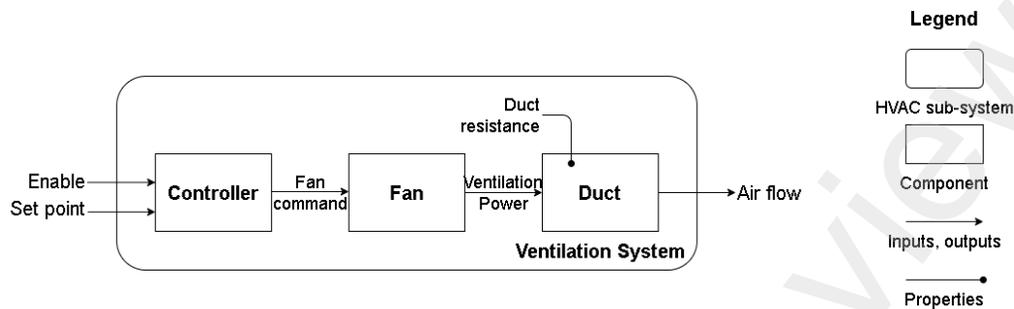


Figure A.8: Control flow model of a ventilation sub-system

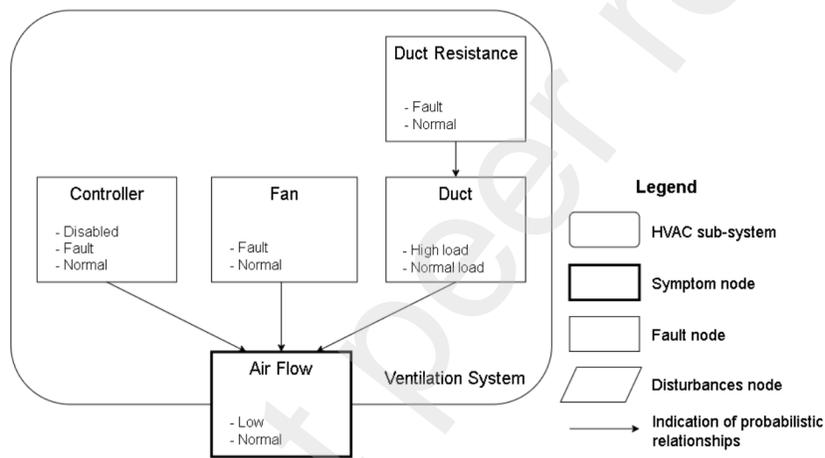


Figure A.9: Fault diagnosis Bayesian sub-network of a ventilation sub-system

water temperature and water flow. The chilled water cooling load is given by the total cooling demand from all served equipment to the system and tank and piping heat loss. The associated fault diagnostic Bayesian sub-network is illustrated in [Figure A.11](#).

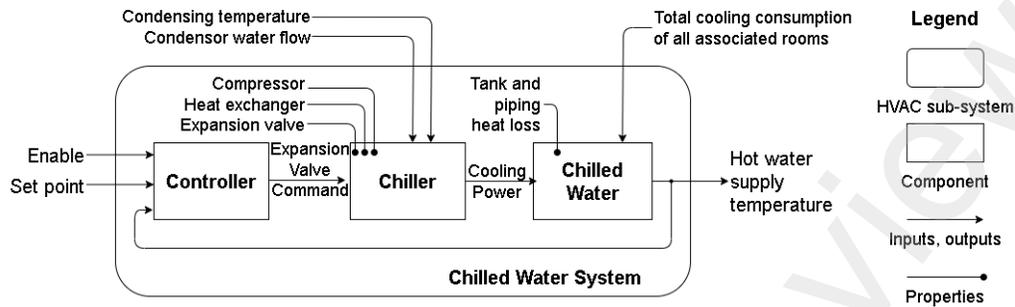


Figure A.10: Control flow model of a chilled water sub-system

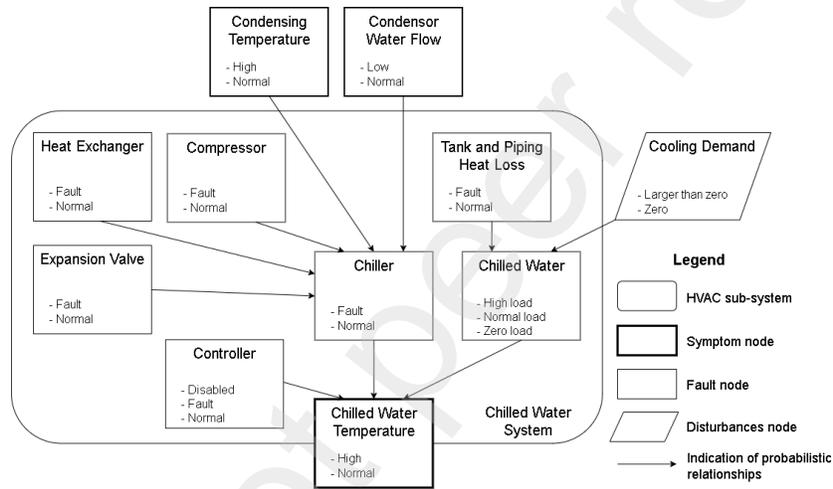


Figure A.11: Fault diagnosis Bayesian sub-network of a chilled water sub-system

Appendix A.6. Summary

Table A.1 summarizes the representation of the systems described previously.

HVAC sub-system	Room Radiator	Hot Water	Hydraulic	Ventilation	Chilled Water
Controlled Variable	Room Temperature	Hot Water Temperature	Water Pressure	Air Pressure	Chilled Water Temperature
HVAC device	Radiator	Boiler	Pump	Air Fan	Chiller
Recipient system	Room	Hot Water	Piping	Duct	Chilled Water
HVAC component1	Valve				Compressor
HVAC component2	Heat exchanger				Heat exchanger
HVAC component3					Expansion
System Characteristics	Building Enclosure	Tank Piping Heat Loss	Piping Resistance	Duct Resistance	Tank Piping Heat Loss
Energy source1	Hot Water Temperature				Condensing Temperature
Energy source2	Hot Water Pressure				Condensing Water Flow
Disturbances	Outdoor air temp.	Heat Demand	Flow Demand		Cool Demand

Table A.1: Summary of HVAC sub-system fault diagnostics Bayesian sub-networks

References

- Ajib, B., Lefteriu, S., Caucheteux, A., Lecoecue, S., 2017. Building thermal modeling using a hybrid system approach, in: 20th World Congress The International Federation of Automatic Control, Toulouse, France. URL: <https://hal.science/hal-01720374>.
- Balaji, B., Bhattacharya, A., Fierro, G., Gao, J., Gluck, J., Hong, D., Johansen, A., Koh, J., Ploennigs, J., Agarwal, Y., Berges, M., Culler, D., Gupta, R., Kjærsgaard, M.B., Srivastava, M., Whitehouse, K., 2016. Brick: Towards a unified metadata schema for buildings, in: Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments, Association for Computing Machinery, New York, NY, USA. p. 41–50. URL: <https://doi.org/10.1145/2993422.2993577>, doi:10.1145/2993422.2993577.
- Barber, D., 2012. Bayesian reasoning and machine learning. Cambridge University Press.
- Chen, Y., Wen, J., Pradhan, O., Lo, L.J., Wu, T., 2022. Using discrete bayesian networks for diagnosing and isolating cross-level faults in hvac systems. Applied Energy 327, 120050. URL: <https://www.sciencedirect.com/science/article/pii/S0306261922013071>, doi:<https://doi.org/10.1016/j.apenergy.2022.120050>.

- European Commission, 2020. A renovation wave for europe - greening our buildings, creating jobs, improving lives. URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0662>.
- Finkelstein, M., 2008. Failure Rate Modelling for Reliability and Risk. Springer Series in Reliability Engineering, Springer London. URL: <https://link.springer.com/book/10.1007/978-1-84800-986-8>.
- Gao, Tianyun, Boguslawski, Bartosz, Marié, Sylvain, Béguery, Patrick, Thebault, Simon, Lecoeuche, Stéphane, 2019. Data mining and data-driven modelling for air handling unit fault detection. E3S Web Conf. 111, 05009. URL: <https://doi.org/10.1051/e3sconf/201911105009>, doi:10.1051/e3sconf/201911105009.
- Gunay, B., Shen, W., Yang, C., 2017. Characterization of a building's operation using automation data: A review and case study. Building and Environment 118, 196–210. URL: <http://linkinghub.elsevier.com/retrieve/pii/S0360132317301336>, doi:10.1016/j.buildenv.2017.03.035.
- Haystack, 2014. About project haystack. URL: <https://project-haystack.org/>.
- International Standards Organization, 2018. ISO 16739-1:2018. Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries — Part 1: Data schema. Standard. International Organization for Standardization. Geneva, CH. URL: <https://www.iso.org/standard/70303.html>.
- Montgomery, R., McDowall, R., 2008. Fundamentals of HVAC control systems. Elsevier.
- Mrad, A.B., Delcroix, V., Piechowiak, S., Leicester, P., Abid, M., 2015. An explication of uncertain evidence in bayesian networks: likelihood evidence and probabilistic evidence. Applied Intelligence 43, 802–824.
- Neapolitan, R.E., et al., 2004. Learning bayesian networks. volume 38. Pearson Prentice Hall Upper Saddle River, NJ.
- Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann.

- Peng, Y., Zhang, S., Pan, R., 2010. Bayesian network reasoning with uncertain evidences. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 18, 539–564.
- Schein, J., Bushby, S., 2006. A hierarchical rule-based fault detection and diagnostic method for HVAC systems. *HVAC&R Research* 12, 111–125. URL: <http://www.tandfonline.com/doi/abs/10.1080/10789669.2006.10391170>, doi:10.1080/10789669.2006.10391170.
- Shi, Z., O'Brien, W., 2019. Development and implementation of automated fault detection and diagnostics for building systems: A review. *Automation in Construction* 104, 215–229. URL: <https://www.sciencedirect.com/science/article/pii/S0926580518312354>, doi:<https://doi.org/10.1016/j.autcon.2019.04.002>.
- Signoret, J.P., Leroy, A., 2021. Reliability assessment of safety and production systems : analysis, modelling, calculations and case studies. Springer series in reliability engineering, Springer, Cham, Switzerland.
- Tungkasthan, A., Jongsawat, N., Poompuang, P., Intarasema, S., Premchaiswadi, W., Jao, C.S., 2010. Automatically building diagnostic bayesian networks from on-line data sources and the smile web-based interface. *Decision Support Systems* , 321–334.
- Turner, W., Staino, A., Basu, B., 2017. Residential HVAC fault detection using a system identification approach. *Energy and Buildings* 151, 1–17. URL: <http://linkinghub.elsevier.com/retrieve/pii/S0378778816316590>, doi:10.1016/j.enbuild.2017.06.008.
- Verbert, K., Babuška, R., De Schutter, B., 2017. Combining knowledge and historical data for system-level fault diagnosis of HVAC systems. *Engineering Applications of Artificial Intelligence* 59, 260–273. URL: <http://linkinghub.elsevier.com/retrieve/pii/S095219761630255X>, doi:10.1016/j.engappai.2016.12.021.