



**HAL**  
open science

## Detection of Faults and Drifts in the Energy Performance of a Building Using Bayesian Networks

David Bigaud, Abdérafi Charki, Antoine Caucheteux, Fally Titikpina, Téodor  
Tiplica

► **To cite this version:**

David Bigaud, Abdérafi Charki, Antoine Caucheteux, Fally Titikpina, Téodor Tiplica. Detection of Faults and Drifts in the Energy Performance of a Building Using Bayesian Networks. Journal of Dynamic Systems, Measurement, and Control, 2019, 141 (10), pp.101011. 10.1115/1.4043922 . hal-02556347

**HAL Id: hal-02556347**

**<https://univ-angers.hal.science/hal-02556347v1>**

Submitted on 8 Nov 2022

**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.

AQ1  
AQ2<sup>1</sup>  
AQ3<sup>2</sup>  
3  
4  
5  
6  
7

**David Bigaud**

LARIS (Laboratoire Angevin de Recherche en  
Ingénierie des Systèmes),  
University of Angers,  
62 Avenue Notre Dame du Lac,  
Angers 49000, France  
e-mail: david.bigaud@univ-angers.fr

**Abderafi Charki<sup>1</sup>**

LARIS (Laboratoire Angevin de Recherche en  
Ingénierie des Systèmes),  
University of Angers,  
62 Avenue Notre Dame du Lac,  
Angers 49000, France  
e-mail: abderafi.charki@univ-angers.fr

**Antoine Caucheteux**

CEREMA (Centre d'Etudes et  
d'expertise sur les Risques,  
l'Environnement, la Mobilité et l'Aménagement),  
23 Avenue de l'Amiral Chauvin BP 20069,  
Les Ponts-de-Cé 49136, France  
e-mail: antoine.Caucheteux@cerema.fr

**Fally Titikpina**

LARIS (Laboratoire Angevin de Recherche en  
Ingénierie des Systèmes),  
University of Angers,  
62 Avenue Notre Dame du Lac,  
Angers 49000, France  
e-mail: fally.titikpina@univ-angers.fr

**Teodor Tiplica**

LARIS (Laboratoire Angevin de Recherche en  
Ingénierie des Systèmes),  
University of Angers,  
62 avenue Notre Dame du Lac,  
Angers 49000, France  
e-mail: teodor.tiplica@univ-angers.fr

# Detection of Faults and Drifts in the Energy Performance of a Building Using Bayesian Networks

*Despite improved commissioning practices, malfunctions or degradation of building systems still contribute to increase up to 20% the energy consumption. During operation and maintenance stage, project and building technical managers need appropriate methods for the detection and diagnosis of faults and drifts of energy performances in order to establish effective preventive maintenance strategies. This paper proposes a hybrid and multilevel fault detections and diagnosis (FDD) tool dedicated to the identification and prioritization of corrective maintenance actions helping to ensure the energy performance of buildings. For this purpose, we use dynamic Bayesian networks (DBN) to monitor the energy consumption and detect malfunctions of building equipment and systems by considering both measured occupancy and the weather conditions (number of persons on site, temperature, relative humidity (RH), etc.). The hybrid FDD approach developed makes possible the use of both measured and simulated data. The training of the Bayesian network for functional operating mode relies on on-site measurements. As far as dysfunctional operating modes are concerned, they rely mainly on knowledge extracted from dynamic thermal analysis simulating various operational faults and drifts. The methodology is applied to a real building and demonstrates the way in which the prioritization of most probable causes can be set for a fault affecting energy performance. The results have been obtained for a variety of simulated situations with faults deliberately injected, such as increase in heating preset temperature and deterioration of the transmission coefficient of the building's glazing. The limitations of the methodology are discussed and are translated in terms of the ability to optimize the experiment design, control period, or threshold adjustment on the control charts used. [DOI: 10.1115/1.4043922]*

*Keywords: energy performance, fault detection and diagnosis, Bayesian network, weather conditions, occupancy*

AQ4

8  
9  
10  
11  
12  
13  
14

15  
16  
17  
18  
19  
20  
21

22  
23  
24  
25  
26  
27  
28

29  
30  
31  
32  
33  
34  
35

**39 1 Introduction**

40 According to different sources, it is commonly considered that  
41 buildings are responsible for about 30–40% of the energy con-  
42 sumption in Western countries. Many studies have demonstrated  
43 that, in the operational and maintenance (O&M) stage, buildings  
44 actually use more energy than estimated by energy simulations in  
45 the design stage [1]. In Ref. [2], a difference of performance of  
46 more than 30% has been estimated. This difference is due to  
47 uncertainty in the modeling of energy simulations [3], as well as  
48 to differences in occupant behavior and in building use over time  
49 (e.g., modifying room functions and building occupancy). Mal-  
50 functions/degradation and a bad control of systems also contribute  
51 to reduction in comfort and increase in energy consumption (in  
52 some cases, 10–20% higher than necessary) [4]. Recently, in Ref.  
53 [5], it has been demonstrated that the averaged cooling energy in  
54 office buildings was about 16% higher than designed due to opera-  
55 tional errors. Even if the efforts are made to improve the continu-  
56 ous commissioning [6], energy consumption is still higher than

57 expected and, therefore, the development of effective preventive  
58 maintenance strategies for building systems is very important.  
59 Thus, condition-based maintenance plans the maintenance accord-  
60 ing to the need determined by the system conditions [7]. Never-  
61 theless, despite examples of preventive maintenance for heating,  
62 ventilation, air conditioning, and refrigeration [8], automated  
63 energy performance diagnosis features are currently scarcely  
64 applied in building energy management systems practices. A key-  
65 step within the condition-based maintenance process is the ability  
66 to monitor the system malfunctioning from the available signals,  
67 hereafter referred to as fault detections and diagnosis (FDD). The  
68 automated FDD tools are useful to alarm and identify faults  
69 promptly and to identify the variables that cause the degradation  
70 of performances, with due regard to the level of accepted risk.

71 Over the last decades, a significant number of researches had  
72 been carried out in the developments of FDD methods for build-  
73 ing's systems (see, e.g., Ref. [9] for air handling units, [10] for  
74 chiller and [11] for HVAC systems). Building energy HVAC  
75 FDD has been proved efficient to reduce energy consumption in  
76 buildings during O&M stages. Some field observations show that  
77 energy savings of 5–30% can be achievable by correcting faults  
78 diagnosed in buildings [12]. Recent FDD case studies in Australia  
79 [13] show that energy savings between 15% and 28% are possible  
80 by implementing HVAC FDD systems. Despite all these benefits,

AQ5

<sup>1</sup>Corresponding author.

Contributed by the Dynamic Systems Division of ASME for publication in the JOURNAL OF DYNAMIC SYSTEMS, MEASUREMENT, AND CONTROL. Manuscript received February 3, 2018; final manuscript received May 27, 2019; published online xx xx, xxxx. Assoc. Editor: Umesh Vaidya.

81 FDD tools are still not broadly used in practice due to the time-  
82 consuming nature of the tasks (which must be repeated for each  
83 building) and to the complexities of FDD algorithms.

84 According to Ref. [14], FDD methods can be categorized into  
85 three types of approaches: quantitative or model-based, qualitative  
86 or rule-based, and process-history-based or data-driven.

87 Quantitative or model-based approaches are usually based on  
88 the knowledge of the physical laws describing the system (e.g.,  
89 heat and mass balances). The common method in the quantitative  
90 FDD approach is the residual generation and analysis (difference  
91 between measured and simulated values). Model-based  
92 approaches do not rely on past data for training; therefore, they  
93 are considered efficient for the detection of unknown abnormal-  
94 ities. They can be used at the component as well as the whole  
95 building levels.

96 Qualitative approaches require a prior knowledge of the system  
97 from which simplified relationships (e.g., the rule-based method)  
98 are developed. This kind of method is rather simple, but it is also  
99 specific to the system under study and often requires a  
100 “customized” implementation, which mainly relies on experts and  
101 developers’ knowledge [15].

102 Process-history-based or data-driven methodologies are also  
103 relatively easy to implement, but they require a significant amount  
104 of data to be efficient. Since they rely on historical data, one has  
105 to study a whole range of various system operations. In addition,  
106 they generally cannot be directly applied on another system with-  
107 out taking into account the specificities of the system [16,17].  
108 Process-history-based approaches include methods such as artificial  
109 neural network (e.g., Ref. [18]), principal component analysis  
110 [19], support vector machine (e.g., Refs. [20] and [21]), Bayesian  
111 classifiers (e.g., Ref. [22]), Fisher discriminant analysis (e.g., Ref.  
112 [23]), etc.

113 In Ref. [24], the authors propose to consider a fourth category  
114 of approaches: hybrid approaches obtained by connecting several  
115 of the aforementioned approaches in order to improve the overall  
116 FDD accuracy and robustness. For example, model-based FDD  
117 approaches are often connected to data-driven methods. The present  
118 research proposes a hybrid approach (data-driven + model-  
119 based). Our purpose is to train the FDD tool based on a large data  
120 collection and completing knowledge through dynamic energy  
121 simulations (DES) in order to study the effect of complementary  
122 unobserved/dysfunctional operations.

123 Bayesian networks (BNs) can be considered as an appropriate  
124 method to accurately represent a building, which is considered as  
125 a complex system with uncertain, incomplete, and conflicting  
126 information. They are deemed a powerful tool to develop expert  
127 systems, and they have the potential to detect and diagnose faults  
128 or drifts. A BN is a probabilistic graphical model that represents  
129 relationships of probabilistic dependence within variables by  
130 means of directed acyclic graphs (see Sec. 2). BNs used as a FDD  
131 tool offer a number of undeniable advantages: the ability to  
132 manipulate continuous and discrete variables, the ability to take  
133 into account time through dynamic Bayesian networks (DBNs),  
134 object modeling via the formalism of object-oriented Bayesian  
135 networks, and the possibility to expand into decision optimization  
136 techniques through the use of influence diagrams (extension of  
137 Bayesian networks) [25]. Beyond being able to identify the causes  
138 of drift or dysfunction, one remarkable aspect of the BNs, that we  
139 cannot find in other methods (such as artificial neural network or  
140 support vector machine), is the ability to sort them from the most  
141 to the least probable and hence to prioritize inspection and mainte-  
142 nance actions. In the frame of research for increasing the energy  
143 performances of buildings, BNs have proved their interest and rel-  
144 evancy, for example, for HVAC [26], with focus on chiller  
145 [22–27], air-handling units [28,29], and heat pumps’ faults  
146 diagnosis [30].

147 In this paper, we develop a generic (in contrast of most other  
148 approaches, actually specific) hybrid FDD approach to detect and  
149 diagnose drifts in energy consumption in a monitored building  
150 instrumented with different sensors. Data are validated and/or

completed by means of energy simulations. Expert knowledge 151  
enhances the hybrid approach by validating the physical meaning 152  
of the architecture of the FDD model. BNs are used here as a 153  
generic model to simulate complex inferences—combining influ- 154  
ences from weather conditions qualified as external loads (outdoor 155  
temperature (OT), relative humidity (RH), wind speed, solar radi- 156  
ation, etc.), variables qualified as internal or intrinsic loads (indoor 157  
air temperature, internal sources, rate of infiltration), influences of 158  
energy systems behavior (heat or cooling production, regulation, 159  
etc.), and the occupants on differing scales of observation (at 160  
“office,” “floor” or “building” levels)—with regard to the level of 161  
energy performance (or consumption) of a building. In practice, 162  
the BN-based hybrid FDD approach proposed in this paper can be 163  
used either for new projects or for existing buildings for which we 164  
aim at enhancing energy efficiency. 165

166 The paper is structured as following: in Sec. 2, the general 166  
aspects of our hybrid FDD approach based on the principle of 167  
Bayesian networks are explained. In Secs. 3 and 4, we present the 168  
case study and the application of the hybrid FDD approach to this 169  
case. In Sec. 5, we propose a deep discussion on the results and 170  
try to show all the abilities of the approach. Finally, we give some 171  
conclusions and discuss about future works in Sec. 6. 172

## 2 Creation of a Bayesian Network to Monitor the 173 Performance of a Building 174

2.1 Principle and Properties of Bayesian Networks. A BN 175  
is a statistical modeling tool whose formalism makes it possible to 176  
deal with uncertainties. The most useful application of BN is to 177  
assess hierarchically the possible causes of risks, failures or opera- 178  
tional drifts [31]. Today, BN can be considered as a key modeling 179  
framework in decision making in a wide variety of domains such 180  
as social sciences [32,33], robotics [34,35], biochemistry or biol- 181  
ogy [36,37], medicine [38,39], engineering [40,41] and, of course, 182  
energy. 183

184 In order to understand fully the properties of BNs, it must first 184  
be clear that they are a hybrid of two different fields: the theory of 185  
graphs and the theory of probabilities. In short, a BN is a graphical 186  
representation of a probabilistic model revealing the different 187  
relationships that the variables of a model can have. It expresses 188  
and factorizes the joint probability of  $m$  variables in  $m$  conditional 189  
independences and its structure enables local calculations of prob- 190  
ability using all the information about the joint distributions. 191  
These conditional independences make it possible to reduce the 192  
number of calculations necessary for the inference and learning of 193  
a probabilistic model by simply reducing the size of its structure. 194  
For example, a joint probability of  $m$  variables is written using the 195  
chain rule (or “general product” rule) as follows: 196

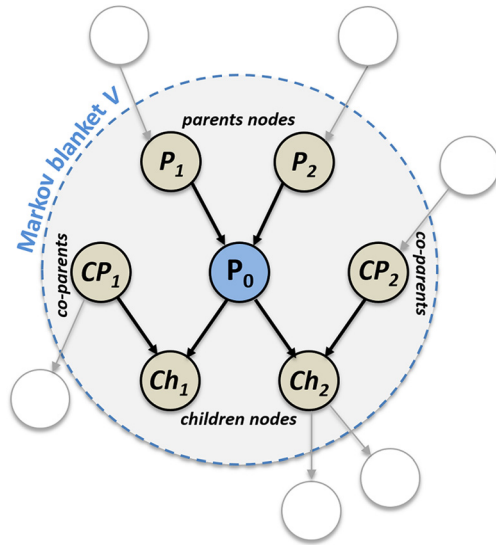
$$p(x_1, x_2, \dots, x_m) = p(x_m | x_{m-1}, \dots, x_1) \dots p(x_2 | x_1) p(x_1) \quad (1)$$

197 This equation can be shortened by introducing or defining the con- 198  
ditional independences between its variables. Moreover, by illus- 199  
trating these independences in the form of a BN, it becomes 200  
possible simply to increase the number of conditional probability 201  
distributions for each variable in accordance with the parents and 202  
rewrite the joint distribution as follows, where  $pa(x_i)$  refers to the 203  
parents of  $x_i$ : 204

$$p(x_1, x_2, \dots, x_m) = \prod_{i=1}^{i=m} p(x_i | pa(x_i)) \quad (2)$$

205 Nevertheless, the fact that the conditional distribution for each 206  
variable is defined according to its parents does not signify that no 207  
other variables influence it. In other words, nodes other than its 208  
parents in the BN can influence a node. These nodes are consistent 209  
with the Markov condition, which states that a variable is isolated 210  
by a subset of variables of the overall set  $V$  known as the Markov 211  
blanket [42]. A variable is therefore conditionally independent of 212





**Fig. 1 The Markov Blanket.** The shaded nodes (parents, co-parents, children nodes) are inside the Markov Blanket of node “A.” The white ones are outside the blanket.

213 other variables outside its blanket if the nodes of the blanket are  
 214 observed. This includes its parent and child nodes and the co-  
 215 parents of its children (see Fig. 1). These nodes, if observed,  
 216 “block” the node in question from other nodes outside its periph-  
 217 ery (see Sec. 4.2 and the discussion related to Fig. 6 for an  
 218 illustration).

219 In the example illustrated in Fig. 1, Eq. (2) is developed as

$$\begin{aligned}
 p(V) &= p(Ch_1, Ch_2, CP_1, CP_2, P_0, P_1, P_2) \\
 &= p(P_0|P_1, P_2) \cdot p(Ch_1|CP_1, P_0) \cdot p(Ch_2|CP_2, P_0) \cdot p(P_1) \\
 &\quad \cdot p(P_2) \cdot p(CP_1) \cdot p(CP_2) \quad (3)
 \end{aligned}$$

220 This property of the BN is not only useful but also essential for  
 222 inference calculation, and makes it possible to determine instanta-  
 223 neously and visually whether a set of variables is conditionally  
 224 independent of another one.

225 The most recent developments in BNs have essentially focused  
 226 on the inference algorithms [43–45] and on the learning of the  
 227 structure and parameters of the network [46]. We will concentrate  
 228 on these aspects in constructing BNs to describe functional and  
 229 dysfunctional modes in the energy systems of a test building.

**230 2.2 Construction Stages of a Bayesian Network for the**

**231 Simulation of a Building.** Building management is increasingly  
 232 relying on automated procedures with sophisticated instrumenta-  
 233 tion on systems and equipment, allowing the collection of consid-  
 234 erable quantities of data. In fact, the volume of data collected is so  
 235 vast that it would be impossible for an operator to monitor directly  
 236 each variable involved in the procedure. It seems perfectly logical,  
 237 therefore, to monitor the technical systems of a building using  
 238 data-driven methods such as the BN.

239 The modeling procedure for a building system comprises three  
 240 stages [47]

- The first stage consists in collecting a database of measured/  
 241 real inputs (climate, envelope, energy systems, occupants),  
 242 and outputs (energy needs) calculated from actual data and/  
 243 or derived from DES. An approach founded entirely on  
 244 measured inputs and outputs is possible when characterizing  
 245 an existing building. For a new building, insofar as the  
 246 energy performance of its equipment and systems cannot be  
 247 measured on long periods, it is very difficult to characterize  
 248 the functional and dysfunctional modes. Therefore, the

hybrid approach, of real inputs and simulated outputs, proves  
 249 to be very effective in the majority of situations. 250

- The second stage consists in learning from the created data-  
 251 base to construct a BN (i.e., to model casual relationships  
 252 and conditional probability distributions between the vari-  
 253 ables that will influence the energy performance) enables to  
 254 mimic the building and its energy systems in their normal  
 255 operational modes. That is to establish a “baseline,” essen-  
 256 tially, from which to observe operational drifts. This stage is  
 257 regarded as the inductive part of the BN’s construction in  
 258 that we use the effects/causes to faults/consequences  
 259 relationships. 259

260 Constructing the BN involves making certain choices. For  
 261 example, continuous nodes may be preferred to discrete ones,  
 262 the choice here being a matter of compromise between complexity  
 263 (and thus calculation time) and precision of the model. Whatever  
 264 the decision made, the robustness of the learning methods of the  
 265 BN needs to be tested and the physical representativeness eval-  
 266 uated by appropriate experts.

- The third stage consists in adapting the model to the dysfunc-  
 267 tional operating modes based on the consolidated architec-  
 268 ture obtained at the previous stage. Here, the detection  
 269 capacity of the BN is tested. In other words, an analysis of its  
 270 sensitivity to operational drifts and consequently its ability to  
 271 detect faults. The dysfunctional operating mode database is  
 272 created from DES by simulating preset multiple input faults  
 273 (see Sec. 4.6 for details). The conditional probability matri-  
 274 ces logically constitute the main output of this so-called  
 275 deductive stage, since the inference rules now relate faults/  
 276 consequences to effects/causes.

277 Once the model is consolidated, it will be operated to sort the  
 278 possible causes of detected energy performance defects into hier-  
 279 archical order [48].

**280 2.3 Modeling of Energy Performance in Functional and**

**281 Dysfunctional Modes.** The modeling the building’s energy per-  
 282 formance in its functional and dysfunctional modes (correspond-  
 283 ing to the second and third stages mentioned previously) is itself  
 284 divided into six steps (detailed in Sec. 4). The first step will con-  
 285 sist in collecting the problem data. The inputs, i.e., the external  
 286 and internal loads factors that influence the energy performance,  
 287 are measured on-site. The outputs (heating needs), however, are  
 288 simulated via TRNSYS (TRNSYS v17/type 56). A calibration sub-step  
 289 had to be carried out in order to adjust the DES results to mea-  
 290 sured data from the test building. The second step will concern the  
 291 choice and use of the accurate algorithm to learn from calibrated  
 292 input and output data values, which have been first discretized in  
 293 order to reduce calculation times. The third step tests the robustness  
 294 of the BN; a process aided by the judgment of experts. The fourth  
 295 step explores the effectiveness and influence of continuous nodes  
 296 rather than discrete ones. The fifth step consists in developing a  
 297 DBN from which we provide updated baselines from calibrated  
 298 data. The sixth and final step is dedicated to the simulation of dys-  
 299 functional modes. We create new databases by means of DES with  
 300 different faults situations from which the DBNs can be updated.

301 Figures 2(a)–2(c) give a conceptual representation of a BN  
 302 modeling functional and dysfunctional modes. Occ\_i (i = 1 to 3),  
 303 Zone\_j (j = 1 to 3), Syst\_k (k = 1 to 3), Sens\_n (n = 1 to 5) repre-  
 304 sent respectively the occupants, zones (or rooms) of the building,  
 305 technical systems, and measurement sensors.

**306 2.4 Control Charts for the Detection of Performance**

**307 Faults and Drifts.** The principle behind the proposed modeling is  
 308 that the fault detection can be viewed as a binary classification  
 309 task: an observed state belongs to either the normal operation  
 310 class (“under statistical control (SC)”) or the faulty operation class  
 311 (“dysfunctional” or “out-of-control”). A  $T^2$  hotelling multivariate  
 312 control chart [49,50] was used. 312

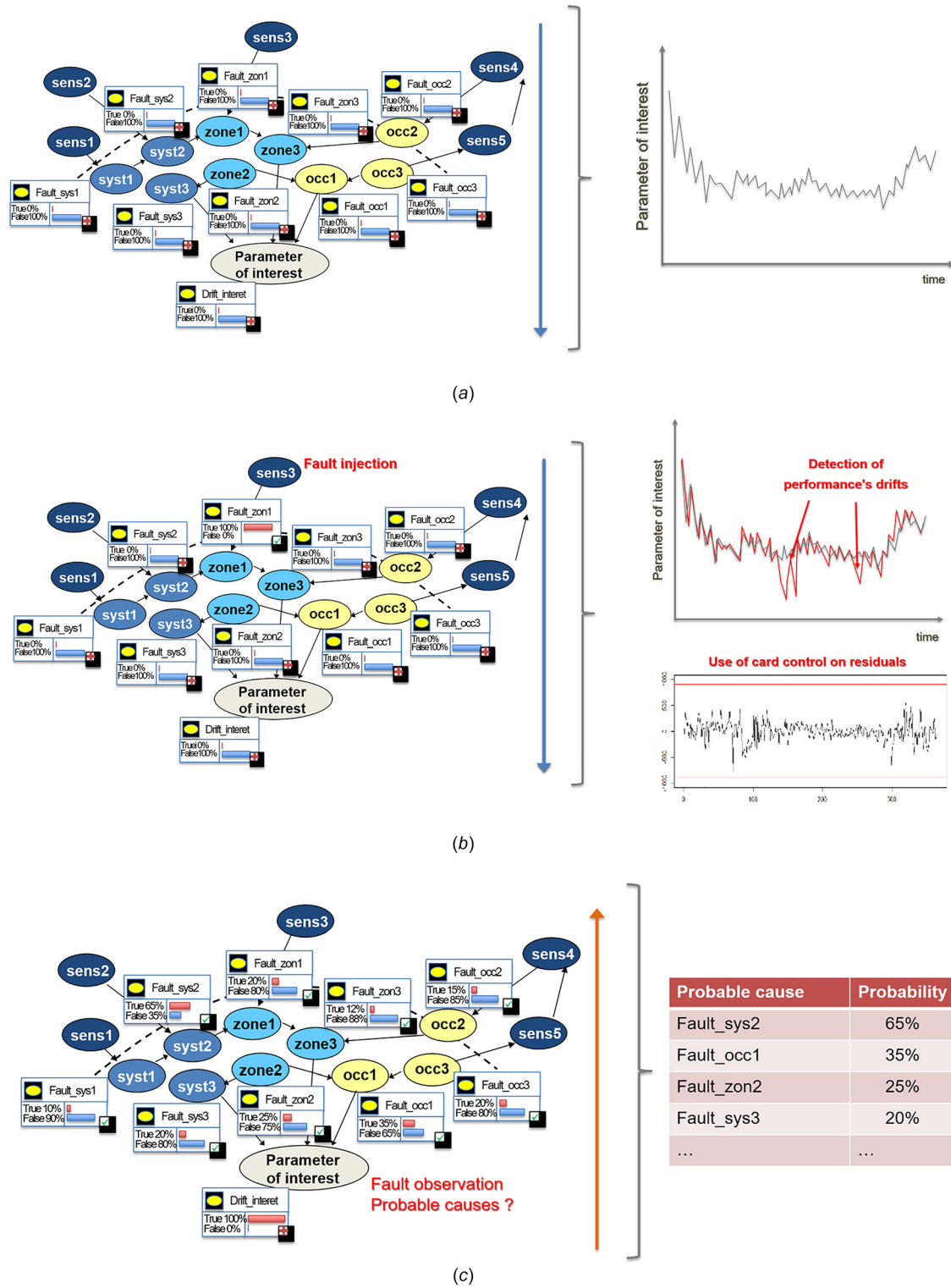


Fig. 2 Three stages of the proposed approach: (a) modeling principle for functional mode—inductive stage to establish “baseline” of behavior in functional (nonfaulty) mode, (b) modeling principle for dysfunctional mode—inductive stage to characterize effects of faults, and (c) modeling principle for dysfunctional mode—deductive stage to identify and hierarchize causes of performance drift

313 The detection of faults (dysfunctional or “out of control” states)  
 314 in the Bayesian network modeling the system can therefore be con-  
 315 sidered by adding extra discrete nodes to the time-monitored  
 316 variables—that means inserting control charts at key points for  
 317 monitoring purposes. In this way, it is possible for example to  
 318 detect abnormally high heating consumption compared with normal  
 319 performance for the considered period, which might suggest a fault.

### 3 Case Study of a Building

320

321 **3.1 Presentation of the Building.** The building we choose  
 322 for our case study belongs to CEREMA in Ponts-de-Cé, in France.  
 323 A surface area of 105 m<sup>2</sup> was instrumented for the purpose of our  
 324 study (see Fig. 3). This building was erected in the 1970s and is  
 325 constituted of common aggregate blocks without insulation.

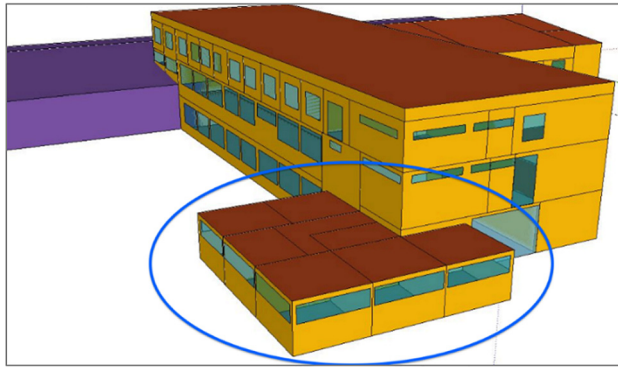


Fig. 3 Modeling using Sketchup 3D® of CEREMA building

326 Insulation was added to the suspended ceilings in the 1990s. The windows  
 327 are double-glazed (4/6/4) with aluminum frames. The building  
 328 uses mechanically controlled single flow ventilation. The heating is  
 329 supplied by a standard high-temperature water loop system.

330 Some hundred sensors were installed to monitor temperature,  
 331 caloric requirements, and occupancy level (see sensors installation  
 332 in Fig. 4). Temperature gauges were placed in each room at a  
 333 height of 1.50 m on interior partition walls, away from windows  
 334 and doors. The caloric requirements were monitored using ultra-  
 335 sonic calorimeters placed at the entry of the heating water loop,  
 336 meaning we may consider that the energy measured is used  
 337 entirely for heating and that heating network losses are moreover  
 338 recovered. In view of its complexity, it was decided to monitor  
 339 the occupancy level using presence (passive infrared motion  
 340 detection sensors) and windows' opening/opening sensors. The  
 341 status of artificial lighting is also checked (with luxometers), and  
 342 the loads due to plug-connected equipment are measured.

343 The TRNSYS tool was used to model the thermal performances of  
 344 the building. The model was calibrated/trained with data collected

345 by Caucheteux et al. [51] during a whole year from Jan. 1, 2013 to  
 346 Dec. 31, 2013. Once trained, we have applied the model for a new  
 347 heating period (151 days between the Oct. 1, 2015 and the Mar. 1,  
 348 2016) in order to assess the accuracy of our thermal model. We  
 349 defined multizone models from the instrumented zones, and the  
 350 Contam tool was coupled to TRNSYS for the airflow modeling (Con-  
 351 tam is a multizone airflow and contaminant transport analysis pro-  
 352 gram. It can help determine airflows such as infiltration,  
 353 exfiltration, and room-to-room airflows in building systems driven  
 354 by mechanical means, wind pressures acting on the exterior of the  
 355 building, and buoyancy effects induced by the indoor and outdoor  
 356 air temperature difference). The model included mechanical venti-  
 357 lation, air infiltration, and all opening of doors and windows. As  
 358 regards air tightness, it was found that the majority of leaks came  
 359 from windows. Therefore, for the purposes of the modeling, infil-  
 360 trations were considered not just as proportional to the surface of  
 361 walls but also linked to the quality of the openings. Finally, the  
 362 occupants were modeled as "persons seated in thermal comfort at  
 363 their workplace." The simulations were carried out to determine  
 364 the caloric requirements in each zone (instrumented + modeled).  
 365 In Figs. 5(a)–5(c), we present comparisons between simulated and  
 366 real measurements for temperatures and required heat quantity  
 367 (HQ). Specifically, Fig. 5(a) presents the variation of the required  
 368 heat quantity as a function of the difference between indoor and  
 369 outdoor temperatures for both simulated and real data. In Figs.  
 370 5(b) and 5(c), we give the details of the variation of the difference  
 371 between indoor and outdoor temperatures and of the required  
 372 daily heat quantity as functions of time (from the Oct. 1, 2015 to  
 373 the Mar. 1, 2016). To assess the agreement between simulations  
 374 and real results, we calculate the normalized mean bias error  
 375 (NMBE) and the coefficient of variation of the root-mean-squared  
 376 error (CV(RMSE)) which are two metrics recommended by both  
 377 ASHRAE and IPMVP [52, 53]. Based on hourly measurements,  
 378 we calculate a NMBE of  $-2.82\%$  and a CV(RMSE) of  $27.02\%$ .  
 379 ASHRAE guideline 14 and IPMVP, respectively, require that  
 380 these values should not exceed 10% and 5% for NMBE, and 30

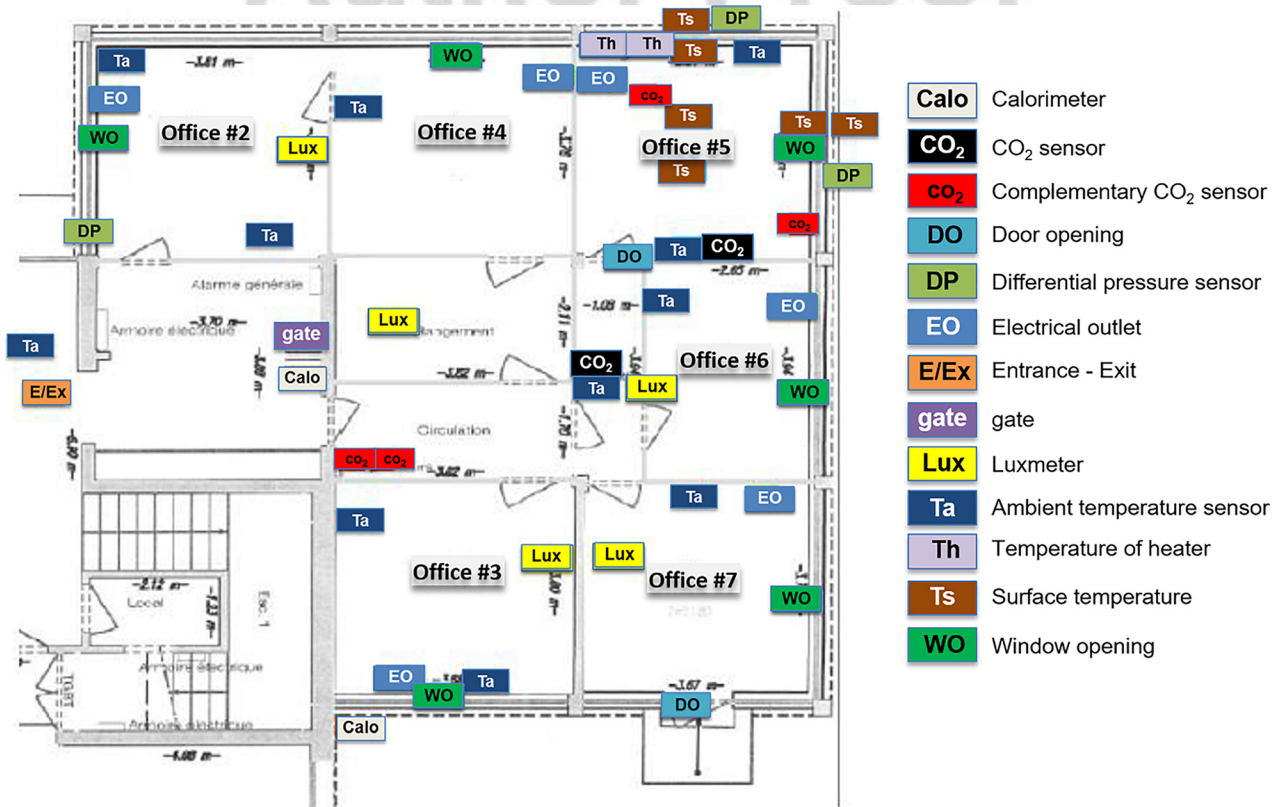


Fig. 4 Positioning of different sensors installed by CEREMA



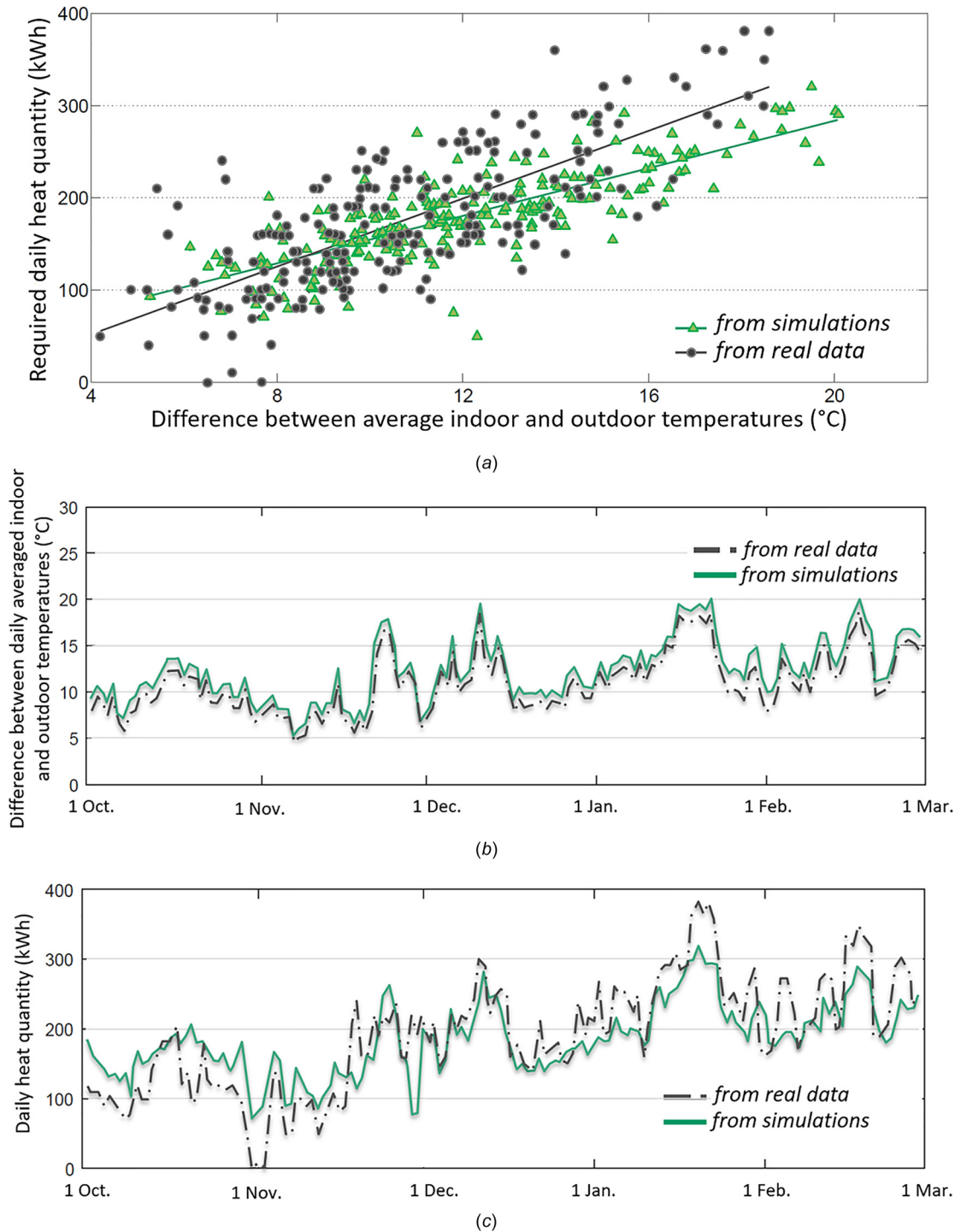


Fig. 5 Comparisons between the results obtained from DES model and actual data: (a) Variation of the required HQ (for the whole surface area of 105 m<sup>2</sup>) as a function of the difference between indoor and OTs for both simulated and actual data, (b) variation of the difference between indoor and OTs as a function of time, and (c) variation of the required daily HQ (for the whole surface area of 105 m<sup>2</sup>) as a function of time

381 and 20% for CV(RMSE). We must note that the CV(RMSE)  
 382 calculated for our DES model is beyond the limit recommended  
 383 by IPMVP. However, since we mainly use the model to detect  
 384 differences between functional and dysfunctional situations and  
 385 not to predict exact energy consumption, we finally consider that  
 386 the DES is accurate in terms of our objective.

**3.2 Characterization of External Loads–Meteorological Data.** All variables of interest related to outdoor physical phenomena influencing the energy performances of the building were considered as external load factors. The first and best known of these phenomena is climatic conditions, which, with year-to-year variability, seasonality effects, and variations over time,

**Table 1 Monitored variables**

Calculated	Measured
Occupancy level	Motion/presence sensors
Interior sources	Entrance and exit of occupants
AL due to windows' opening <sup>a</sup>	Opening/closing of windows and blinds
	Intensity of light
	Concentration of CO <sub>2</sub>
	Electrical energy
	Indoor/preset temperature
	Surface temperature
	Supply air temperature
	Pressure differential
	Heat energy consumption

<sup>a</sup>This variable is deduced from the measurements but is not considered as an internal load. It is used along with the permeability, wind velocity, and relative pressure data to characterize the leakages.

393 determine the performance level required for a building and its  
 394 equipment to ensure the comfort of its occupants.

395 In our study, dynamic meteorological data were collected via a  
 396 weather station installed on-site. This was equipped with the tem-  
 397 perature, relative humidity, global horizontal radiation, wind  
 398 speed, and direction sensors.

399 The station location must be chosen in such a way that it com-  
 400 ply with WMO recommendations [54], i.e., away from obstacles  
 401 and in a grassy area. Wind speed measurements may sometimes  
 402 be affected by certain external phenomena. The measurements  
 403 taken were therefore broken down into two: light winds and strong  
 404 winds. If the anemometer has a resolution of 1 m/s, only wind  
 405 speeds above 1 m/s can be measured.

406 To complete the meteorological data, the renewal of air was  
 407 monitored through the ventilation flow rates and the air perme-  
 408 ability of the building. The flow-rate measurement of the single  
 409 flow ventilation installed was achieved via by an air flow-rate  
 410 measurement cone connected to a hot wire sensor or anemometer.

411 **3.3 Characterization of Internal Loads.** The behavior of  
 412 building occupants can make it difficult to follow the conventional  
 413 scenarios outlined by the regulations in place. People interact with  
 414 their immediate environment searching for desired comfort level  
 415 of which the most direct indicator is indoor temperature. In adjust-  
 416 ing this temperature (or other systems), occupants influence the  
 417 energy consumption.

418 Electrical equipment and appliances are supplementary sources  
 419 of energy. These variables—indoor or preset temperature, number  
 420 of occupants, internal sources—contribute to the building dynam-  
 421 ics and have to be considered as internal loads. These interdepend-  
 422 ent variables are indirectly characterized via the installed sensors.  
 423 Table 1 lists all variables that are directly measured or (indirectly)  
 424 calculated.

425 Occupancy can become a significant internal load. By  
 426 “occupancy” is meant the number of persons present in the build-  
 427 ing at a given time in conjunction with their behavior. The energy  
 428 consumption of a building is directly linked to the occupancy rate  
 429 as well as the behavior of the users. Workplace occupants can be  
 430 a serious source of loss and this is a factor that affects the building  
 431 dynamics: for instance, opening doors and windows necessitates  
 432 an additional supply of heat energy in order to balance thermal  
 433 losses. However, occupants can also be a source of energy by their  
 434 activity. In our study, we have considered that every occupants  
 435 are sitting persons in thermal comfort. Their metabolic rate is  
 436 about 58 W/m<sup>2</sup>/person (1 Met) [55] with a body surface for aver-  
 437 age adult of 1.7 m<sup>2</sup>; therefore, the energy released by metabolism  
 438 is taken at 100 W/person (direct energy source). Devices used by  
 439 occupant (household appliances in a residential building and IT-  
 440 related in a tertiary building) also supply energy to the building  
 441 that can be quantified as an internal source (indirect energy  
 442 source). This indirect energy source is dissociated from the occu-  
 443 pancy and dealt with independently in the form of internal  
 444 sources.

445 To estimate the occupancy level of a building, several estima-  
 446 tors are used based on different measurement protocols that have  
 447 been established [51–56]. They can be based on video recordings,

measurement of CO<sub>2</sub> concentration, motion detection, noise anal- 448  
 ysis, or electricity consumption data [57,58]. The combination of 449  
 several of these measurements provides more detailed information 450  
 regarding the occupancy level and thus a reduction of the uncer- 451  
 tainty of the estimation [59]. In our study, the occupancy level is 452  
 characterized via measurements from motion detection sensors 453  
 (passive infrared), CO<sub>2</sub> sensors, lux meters for artificial lighting, 454  
 sensors for opening/closing of windows, and electricity consump- 455  
 tion sensors for socket loads. As far as internal heat sources are 456  
 concerned, it is quite a complex matter to measure the quantity of 457  
 heat effectively released by occupants and electrical equipment. 458  
 The most common practice consists in measuring electricity con- 459  
 sumption and assuming that all the power consumed by electrical 460  
 equipment is totally released within the building. 461

The indoor temperature is the ambient temperature of a room 462  
 that is regulated via a preset temperature in order to achieve the 463  
 desired thermal comfort. The indoor temperature can be measured 464  
 in every room. In addition, surface temperature measurements can 465  
 be taken, as well as measurements of the air temperature at the 466  
 exit of the heating supply devices. 467

**4 Construction of the Bayesian Network 468**

The construction of the BN follows the six steps mentioned in 469  
 the previous Sec. 2.3. 470

**4.1 Step 1—Collection of Data and Prediction of Heat 471**  
**Quantities by Means of Dynamic Energy Simulations. The 472**  
 inputs and outputs of the model are shown in Table 2. 473

The indoor data were collected from a floor of the building with 474  
 sensors installed according to the measurement plan shown in 475  
 Fig. 4. The outdoor data were mainly obtained from our meteor- 476  
 ological station and completed with data from a regional station. 477  
 All the data were recorded from Jan. 1, 2013 to Dec. 31, 2013. 478  
 The outputs are the predicted heat quantities in each room of the 479  
 building; predictions were made through the DES model. 480

**4.2 Step 2—Choice of the Learning Algorithm for the 481**  
**Construction of the Bayesian Networks. The construction of a 482**

BN consists mainly in finding the set of relationships (links) 483  
 between variables (nodes) and the conditional probabilities tables 484  
 between the thus linked variables. The goal here is to find the BN, 485  
 with a controlled level of complexity, which best fits the training 486  
 database and can give accurate predictions. We typically have to 487  
 deal with a complex combinatorial problem, which first requires 488  
 specific algorithms to detect conditional independences and then 489  
 heuristic search in the solutions space. We have tested several cur- 490  
 rent learning algorithms to find the one(s) that are the most appro- 491  
 priate for our problem. Table 3 shows all the different alternatives 492  
 we have compared. Most of these inference algorithms dedicated 493  
 to discrete nodes are available in the HUGIN software or toolboxes 494  
 like BNT computed with MATLAB by Murphy [60]. 495

From the same database, an algorithm alone can provide multi- 496  
 ple nearly optimal solutions, and, another algorithm can give dif- 497  
 ferent candidate solutions. A way to compare the efficiency of the 498  
 algorithms is to calculate the biased variances for each model 499

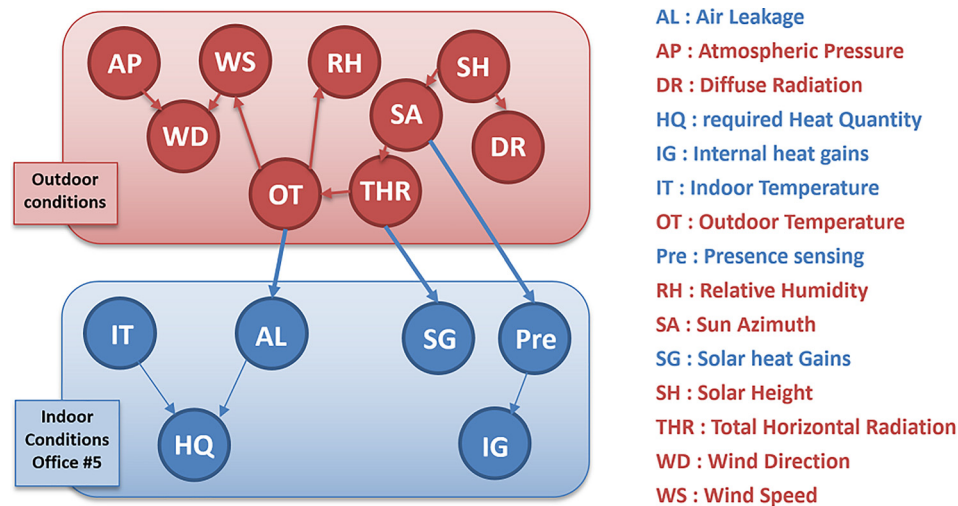


**Table 2 Input and output variables of the model constructed by BN**

Inputs (collected from Jan. 1, 2013 to Dec. 31, 2013)		Output (estimated from DES simulations)
Outdoor—meteorological data	Indoor—measured directly from sensors or calculated	Required HQ calculated individually for each room
Atmospheric pressure (AP) Diffuse radiation (DR) OT RH Solar Azimuth (SA) Solar height (SH) Total horizontal radiation (THR) Wind direction (WD) Wind speed (WS)	AL Internal heat gains (IG) Indoor temperature (IT) Presence or occupancy level (Pre) Solar heat gains (SG)	

**Table 3 The different combined or hybrid learning algorithms**

Search for conditional independence relationships	Hybrid algorithms
Grow-shrink Incremental association	Restricted maximization Max-min hill-climbing Max-min parents and children search
Heuristic search in solution space	Hiton parents and children search
Greedy search Tabu search	Chow-Liu Aracne search



**Fig. 6 Bayesian network for operational mode solely for Office #5 (with hourly measurements over a period of a year)**

500 using an information criterion for the selection of the model (e.g.,  
 501 akaike information criterion (AIC)—or Bayesian information criterion  
 502 (BIC)); the main concern here is to use the same score  
 503 when comparing the BN architectures in order to choose the one  
 504 which is the best in an objective way.  
 505 In order to test our method, we applied it to a single office (see  
 506 Fig. 4—corner office #5) for which a large range of sensors and  
 507 data were available.  
 508 Figure 6 shows the inputs (parent nodes) taken from meteorological  
 509 data (upper part of the diagram) and those taken from  
 510 measurements obtained from sensors inside the rooms (lower  
 511 part). This figure reports one of the best candidate solution  
 512 obtained, considering hourly measurements over the whole year,  
 513 with nodes discretized via an algorithm based on AIC. The marginal  
 514 probabilities and tables of conditional probabilities for the nodes  
 515 were learned from a no-fault data file (since we are in operating  
 516 mode).

It needs to be specified clearly that the purpose of our methodology  
 is not to establish automatically the final BN architecture but just  
 to give to the user a starting point reflecting visually the causal  
 relationships between the various variables. Experts from a physical  
 understanding of various phenomena must validate these conditional  
 dependences and independences. Their expertise is essential in  
 the construction of the final BN architecture. Moreover, the  
 experts may add or delete some arcs in the network if they consider  
 that necessary.  
 As an example of causal relationships explanation, the network in  
 Fig. 6 shows that the energy need (output of the child node) HQ,  
 equivalent to the heating consumption, is directly dependent on  
 indoor temperature IT but equally on the air leakage (AL) level of  
 the envelope. However, the fact that the conditional distribution  
 for each variable is defined according to its parents does not  
 signify that no other variables influence it. Indeed, nodes IG and  
 HQ are conditionally independent if and only if all the nodes

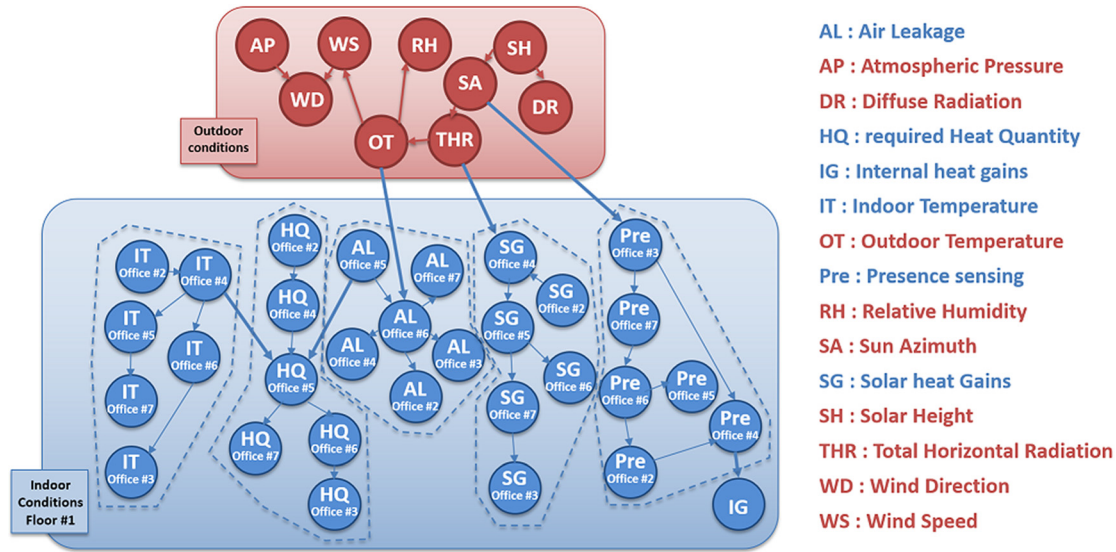


Fig. 7 Bayesian network for operational mode for whole floor (with hourly measurements over a period of a year)

534 belonging to the Markov blanket of IG (or HQ) are observed (i.e.,  
 535 one knows their values). Thus, in Fig. 6, HQ is conditionally  
 536 independent of IG if the nodes IT and AL are observed. That  
 537 means that knowing IG when IT and AL are observed does not  
 538 change anything in the computation of the probabilities of the various  
 539 modalities of the node HQ. Nevertheless, if AL is not  
 540 observed, then IG influences the probabilities of the modalities of  
 541 the node HQ because of the path: IG-Pre-SA-THR-OT-AL-HQ.  
 542 In addition, the OT node influences the HQ if AL is not observed.  
 543 From a physical point of view, if the AL is inexistent (perfect air-  
 544 sealing quality) the OT should not influence the HQ (i.e., HQ is  
 545 influenced by OT only if AL exists).

546 The same task of modeling the operational mode by BN was  
 547 carried out for a whole floor by “forcing,” during the network  
 548 architectural stage, the grouping together of variables of the same-  
 549 type (for example, heating consumption HQ for offices nos. 2–7  
 550 combined). The relationships between nodes of the same type are  
 551 shown with thin-lined arrows. The relationships between nodes of  
 552 different types (for example, OT and air leakage in the office #6  
 553 AL\_Office#6) are shown with thick-lined arrows. Of course,  
 554 many number of different BN models for operational mode are  
 555 possible, limited only by the extent to which the physical coher-  
 556 ence of the relationships established between the network varia-  
 557 bles can be justified. In Fig. 7, we show the BN adopted based  
 558 upon a minimized BIC index, from which we can discuss some  
 559 examples of causal relationships. Physically, there is a link  
 560 between outdoor (OT, RH) and indoor (indoor temperature (IT))  
 561 conditions. The network shown in Fig. 7 proposes a link between  
 562 RH and the AL. It is not what was first expected, but the link is con-  
 563 sistent. Indeed, considering that AL can be regarded as a heat loss  
 564 by air leakage, there is an obvious link with the outdoor conditions.  
 565 Some links are less intuitive. For example, the link between Solar  
 566 Azimuth (SA) and people presence (Pre) was unexpected. Never-  
 567 theless, if we consider the correlation between SA and the hour of  
 568 the day, it is now obvious to see a relationship between the hour of  
 569 the day and the presence of people. Of course, all relationships can  
 570 be discussed but a coherent explanation can be provided for most  
 571 of them and the BN can be consider as acceptable.

572 **4.3 Step 3—Study of the Bayesian Networks Robustness**  
 573 **Against Database Reduction.** The ability of a BN to detect and  
 574 diagnose faults is conditional upon its robustness against param-  
 575 eters related to the data exploitation or extraction. We need to  
 576 check if the general architecture of the BN and its capacity to  
 577 model the data is deeply modified when we consider only a part of

the database or when we take the average of the hourly measure-  
 578 ments on a daily base. The main objective is to see eventually if  
 579 we can construct a BN with equivalent quality but with fewer data  
 580 to process.  
 581

582 First, keeping to the same hourly measurements, we have  
 583 reduced the database to data for which heating consumption was  
 584 different from zero for all offices. We thus excluded data for  
 585 which an energy need had not been calculated for at least one of  
 586 the six offices. The new BN obtained in this case is presented in  
 587 Fig. 8. Comparison to the equivalent BN with the complete data-  
 588 base shows only minor changes in the relationships between varia-  
 589 bles of the same type and in the relationships between types. Only  
 590 one significant relationship appears (see black thick arrow),  
 591 between the indoor temperature in office #6, IT\_office#6, and the  
 592 energy need in this office HQ\_office#6. Therefore, this first modi-  
 593 fication of the quantity of data taken into account and the very  
 594 slight change of the BN seems to demonstrate the robustness of  
 595 our method and that we can reduce the database to the periods  
 596 recording energy needs.

597 The second test consists in modifying the frequency of data  
 598 acquisition. Here, we examined the sensitivity of the BN to a  
 599 change from hourly to daily measurements; the inputs are the  
 600 averages of the values recorded over a day for all characteristics  
 601 measured (including, e.g., wind direction, in degrees). In Fig. 9,  
 602 we show the new BN modeling the relationships between varia-  
 603 bles for the whole floor #1 reducing data to daily measurements  
 604 over a period of a year. The results are interesting: although a  
 605 large majority of the internal relationships between nodes remains  
 606 after change of frequency of measurement, one can see a higher  
 607 number of new links between characteristics of different types  
 608 (see black thick arrows).

609 The two BN’s robustness tests we have conducted allow us to  
 610 conclude that we can use a reduced part of averaged data with  
 611 small disturbances of the BN architecture and of its efficiency.  
 612 The calculation time decreases which will be useful when we will  
 613 adopt continuous nodes rather than discrete ones in step 4 (Sec.  
 614 4.4) and when we will need to develop DBNs in step 5 (Sec. 4.5).

615 **4.4 Step 4—Choice of Type of Nodes.** In the following  
 616 steps, having confirmed the effectiveness of the construction  
 617 method of the BNs for the purposes our study, we choose to not  
 618 discretize the nodes in order to avoid any loss of information; all  
 619 the nodes will be modeled from continuous statistical distribu-  
 620 tions. We maintained the assumption that they follow a Gaussian  
 621 distribution law and that they are not interlinked via linear

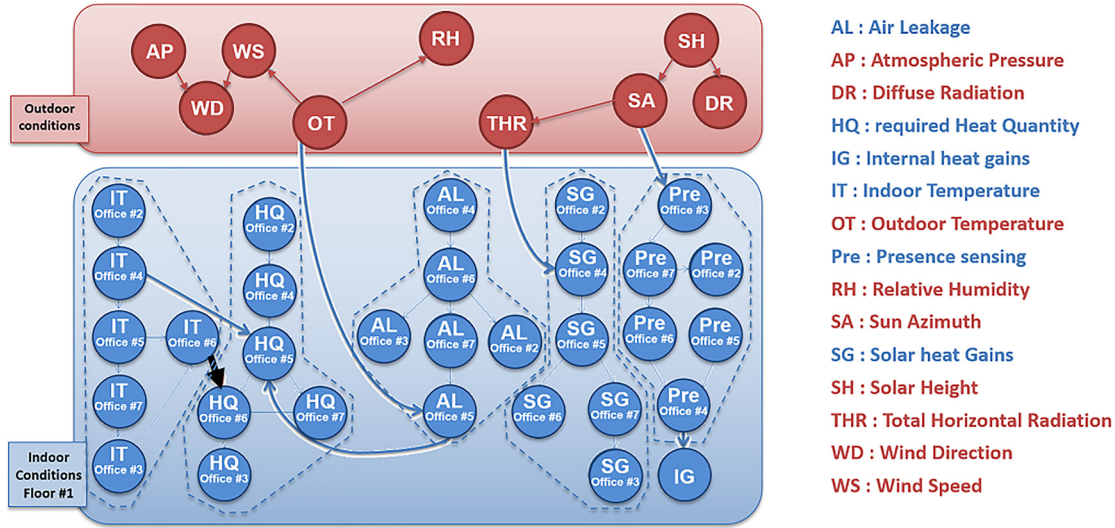


Fig. 8 Bayesian network for operational mode for whole floor (using hourly measurements during period of heating only)

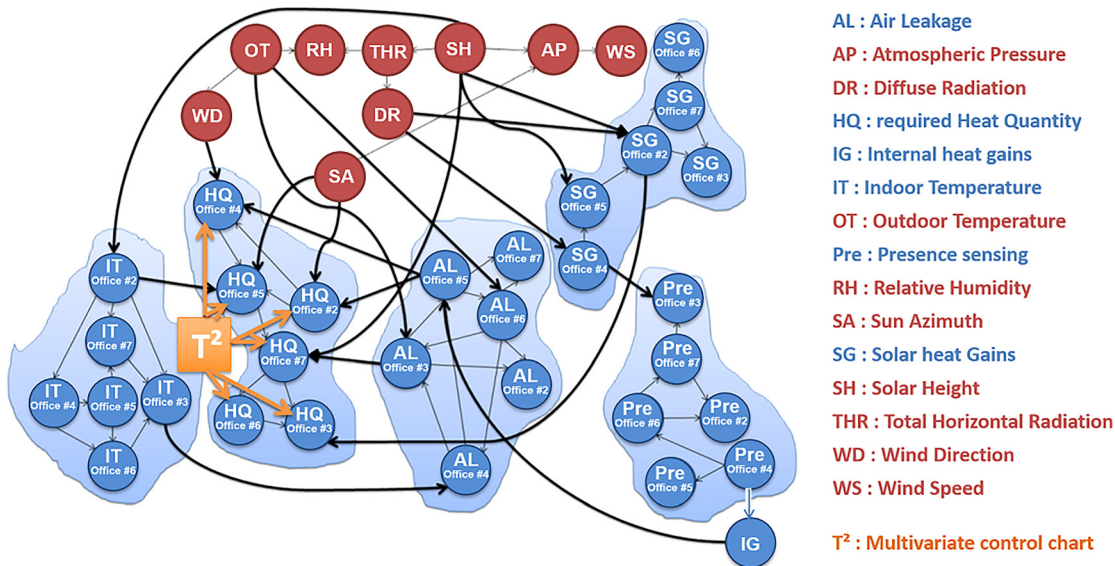


Fig. 9 Bayesian network for whole floor (using daily measurements over a period of a year). Implementation of a  $T^2$  Hotelling multivariate control chart to detect drifts of the energy needs is also illustrated.

622 dependency relationships. It should be noted that these assumptions  
 623 can change depending on whether one deals with a discrete  
 624 or a specific inference node (nonlinear non-Gaussian).

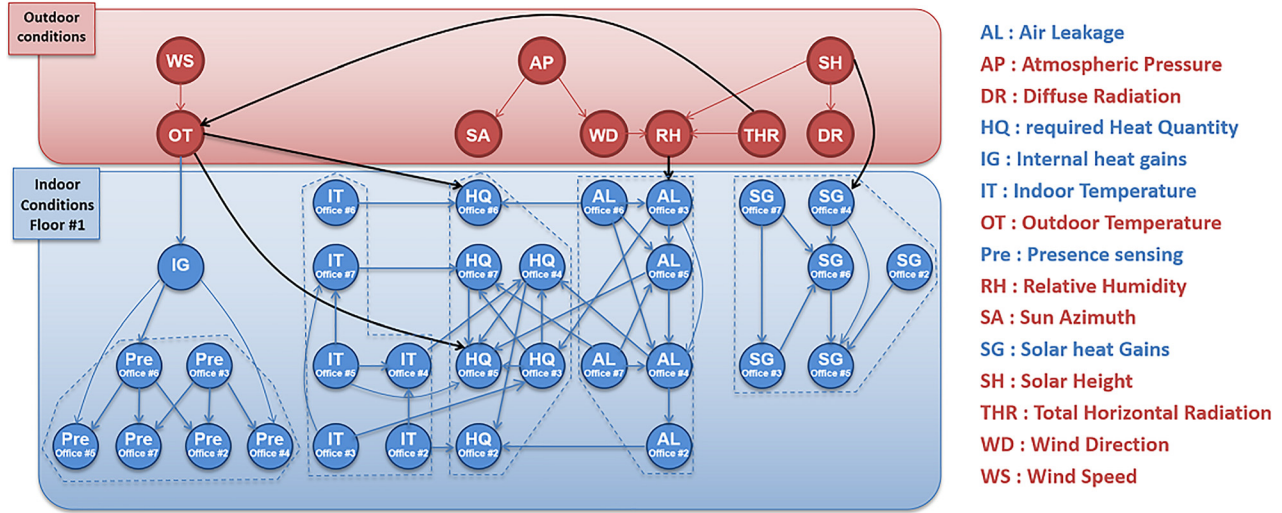
625 Once the network structure and its parameters are defined, it is  
 626 essential to use propagation algorithms that are appropriate to the  
 627 chosen structure in order to calculate marginal distributions as  
 628 well as probability distributions for each variable. Several inference  
 629 algorithms have been suggested in the literature for Gaussian  
 630 conditional networks. We used two algorithms based on the junction  
 631 tree [61,62]. We implemented them with the *R* scientific computing  
 632 environment. These algorithms allow accurate inference in  
 633 Gaussian conditional networks.

634 We also wanted to guarantee better AIC or BIC scores (information  
 635 criterion) for the choice of BN models. It is therefore necessary  
 636 to reduce the impact of local optima during optimization of the  
 637 architectures and parameters of the BNs. We thus adopted the  
 638 bootstrap resampling method already used in Sachs et al. [63]. In  
 639 practical terms, this consists in repeating the learning of the structure  
 640 several times, which allowed us to explore a large number of

641 networks, to average the networks obtained (see Ref. [64]) and to  
 642 finally conserve only the relationships (links) that were present in  
 643 at least 85% of the networks. Figure 10 illustrates the final BN  
 644 obtained with continuous nodes. Some changes are observed  
 645 between this BN and the one obtained with discrete nodes, but the  
 646 general architecture remains. The BIC score for this BN with continuous  
 647 nodes is nearly the same that for discrete nodes ( $-49,727$   
 648 versus  $-50,523$ , i.e., a sensible loss of 1.6%). We will use continuous  
 649 nodes for the upcoming simulations with the aim to obtain  
 650 results that are more accurate.

651 **4.5 Step 5—Creation of Bayesian Network in Dynamic Mode.**  
 652 The DBNs are a special class of BNs; it includes the effect  
 653 of time by considering that the value of a variable at a given time  
 654 can influence its own value at the next time [65]. If we consider a  
 655 set of  $n$  variables  $\mathbf{D}(t) = \{D_1(t); D_2(t); \dots; D_n(t)\}$  varying with  
 656 time, a DBN adds the joint probability distribution of these variables  
 657 for a bounded interval  $[0; T]$ . Generally, this distribution can





AL : Air Leakage  
 AP : Atmospheric Pressure  
 DR : Diffuse Radiation  
 HQ : required Heat Quantity  
 IG : Internal heat gains  
 IT : Indoor Temperature  
 OT : Outdoor Temperature  
 Pre : Presence sensing  
 RH : Relative Humidity  
 SA : Sun Azimuth  
 SG : Solar heat Gains  
 SH : Solar Height  
 THR : Total Horizontal Radiation  
 WD : Wind Direction  
 WS : Wind Speed

Fig. 10 The Bayesian network for whole floor (using daily measurements over a period of a year) with continuous nodes

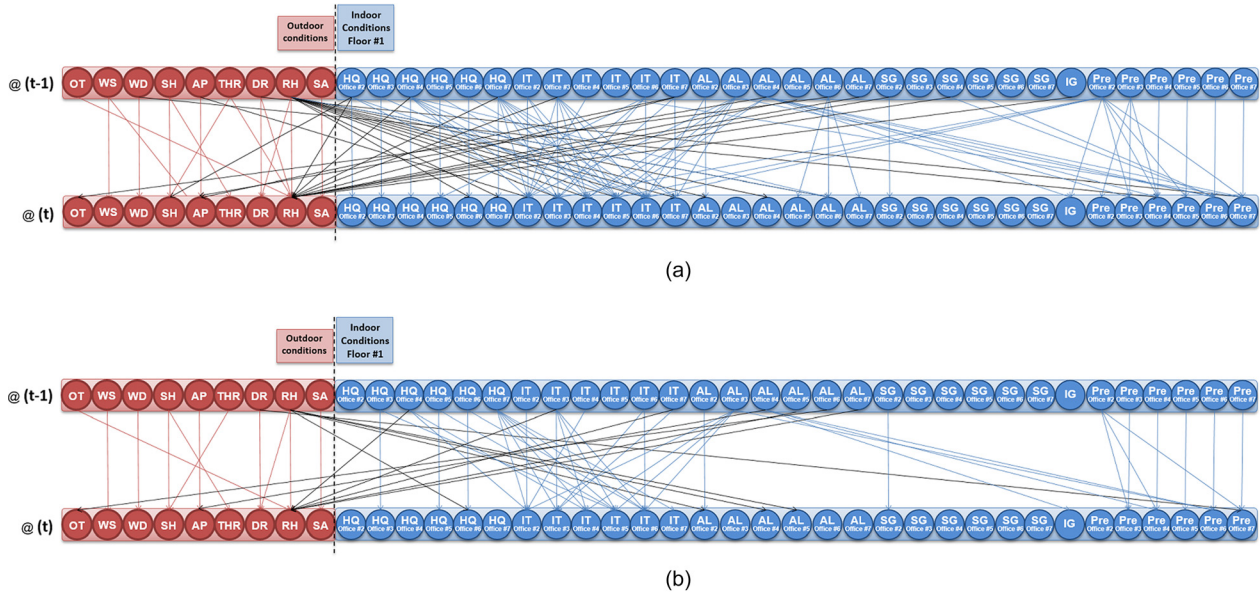


Fig. 11 Illustrations of a dynamic Bayesian network with two different levels of significance: (a) Relationships between variables at  $(t - 1)$  and  $(t)$  for a level of significance  $p$ -value  $< 0.01$  and (b) relationships between variables at  $(t - 1)$  and  $(t)$  for a level of significance  $p$ -value  $< 0.001$

658 be expressed as a static BN with  $T \times n$  variables where  $T$  is the  
 659 number of time intervals considered. If the process we consider to  
 660 model is stationary, the assumptions of independence and the  
 661 associated conditional probabilities are identical for all time intervals  
 662  $\Delta t = (t) - (t - 1)$ . In this case, the DBN can be represented by  
 663 a BN whose structure is duplicated for each time-step. A node  
 664 therefore represents a random variable whose value indicates its  
 665 state at time  $t$ .  
 666 The DBN are widely studied by researchers. In the domain of  
 667 energy optimization, DBN are used as a prognostic approach for  
 668 modeling aggregated load of HVAC systems [66,67], for detecting  
 669 occupant's interactions with windows [68] or for evaluating the  
 670 reliability of grid-connected photovoltaic systems [69]. In our  
 671 case, the prognostic capacity of DBNs is an important improvement  
 672 in the anticipated detection of operational faults and drifts.  
 673 Given that the number of dimensions rises as a result of the intro-  
 674 duction of the time factor, the problem of optimization related to  
 675 this choice of BN model may be dealt with via the "least absolute  
 676 shrinkage and selection operator" or LASSO algorithm explained

in [70]. Interested readers may use the "lars" package for  $R$  software  
 in order to make computations.  
 Figures 11(a) and 11(b) present the (repeatable) relationships  
 between the variables of the BN at times  $(t - 1)$  and  $(t)$  for two  
 levels of significance  $p$ . This  $p$ -value corresponds to the probability  
 that the hypothesis of an existing relationship between two variables  
 at two consecutive times is null. In the following simulations,  
 we have only considered the relationships with a  $p$ -value of 0.001,  
 which in addition contributes to reduce the complexity of the  
 DBN and the simulations time.  
 Figure 12 illustrates how a DBN can be used to predict energy  
 needs or the heat quantity in office #6 (HQ\_office#6) using relationships  
 of dependence between time  $t$  and  $t - 1$ . The curve with the  
 continuous orange line shows the results of simulation by DBN  
 for the predicted/probable values of heat quantity HQ\_office#6  
 with time (for a time sequence  $t$ ), considering the values for  
 all other variables at instants  $t - 1$  and  $t$ . The inputs are daily  
 measurements, and the curve shows energy needs over a year. The  
 dashed blue line on the same figure gives real heat quantity

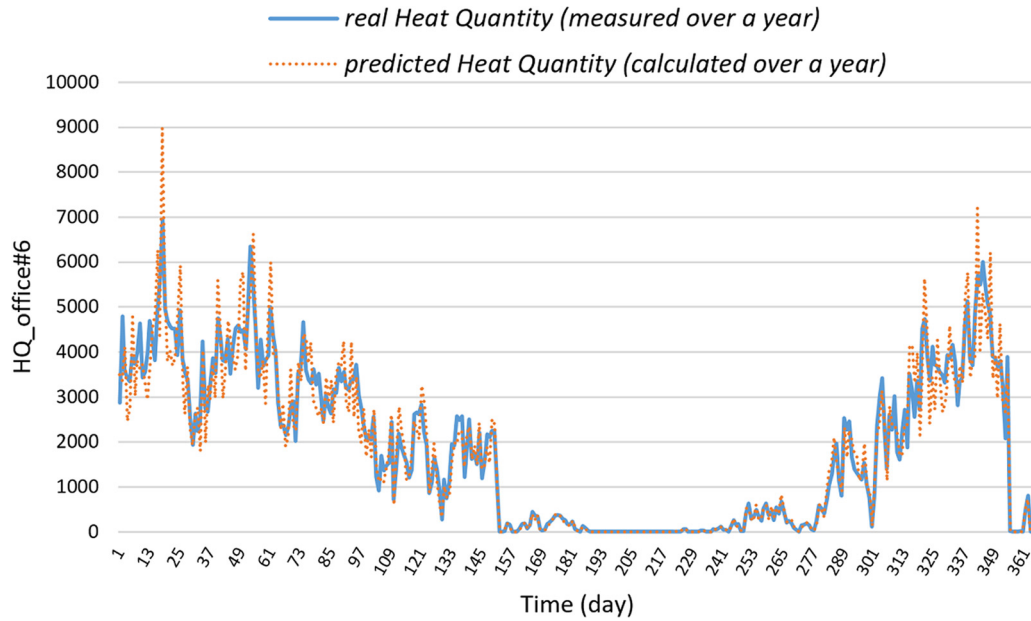


Fig. 12 Energy needs (or HQ) of office #6 (HQ\_Office#6)

696 HQ\_office#6 measured over the year. The fit between the two  
 697 curves is very good as long as the target variable HQ\_office#6 is  
 698 calculated with all the available information (including the heat  
 699 quantities of other offices, HQ\_office#j). In terms of metrics of  
 700 comparison between the actual and simulated heat quantities, we  
 701 found values of NMBE and CV(RMSE) respectively at  $-3.8\%$   
 702 and  $18.7\%$  which are still below the threshold values recom-  
 703 mended by ASHRAE Guideline [52] and IPMVP [53]. For com-  
 704 plete information, the Frechet distance between actual and  
 705 predicted values is estimated at 2703. Regarding more precisely  
 706 the fit of the two curves, it is worth pointing out that the simu-  
 707 lations give a lot of noise for the “warm” period, which, however,  
 708 is of no consequence in the context of assessment of the overall  
 709 performances guarantee. Therefore, we can estimate that the use of  
 710 DBNs for this inductive stage to setup a “baseline” (see Fig. 2(b)),  
 711 from which to detect faults and drifts, is relevant.

712 **4.6 Step 6: Experiment Design Used for Simulating the**  
 713 **Effects of Fault Situations.** Once the baseline corresponding to  
 714 the normal operating conditions has been obtained, the following  
 715 step is used to simulate the effects of small faults on the dynamic  
 716 performances of the building. Then, we create a new database by  
 717 means of DES with different faults situations from which the  
 718 DBNs can be updated. It is important to underline that, for this  
 719 step, we consider that the architecture of the previous DBNs  
 720 obtained in step 5 is stable; the updating will only concern the  
 721 conditional probability tables. This makes the assumption that  
 722 the faults effects, of which amplitudes still remain controlled at  
 723 the (early) time of detection, had a priori been modeled by the pre-  
 724 vious DBNs, considering the whole ranges of the variables values  
 725 of the learning database that encompass the faults amplitudes.

726 From the 14 input variables (the 15th variable is the output heat  
 727 quantity) constituting the basis of the DBN, only a part of them  
 728 are directly requested as factors for the dynamic energy simu-  
 729 lations. All other variables depend on additional factors. Table 4  
 730 summarizes the 13 factors for which we have studied the effects  
 731 of possible deviations from normal conditions. The fourth first  
 732 factors are variables from the DBN, the nine others are factors on  
 733 which depend the remaining ten variables of the DBN. In our  
 734 DESs, the deviation could be either positive or negative depend-  
 735 ing on whether the fault corresponds to a deficit or an excess (i.e.,  
 736 to a high or low performance threshold). In order to update the  
 737 DBN, 104 multiple-fault situations have been simulated by DES.

These situations are summed up in Table 10 in Appendix; all the  
 assumptions and data shown in this table are reported in Ref. [56].

## 5 Simulation and Results

5.1 **Results for Certain Simulated Faults.** Once the 104  
 fault situations are simulated by DES, we have studied various  
 fault situations in order to test the FDD capabilities of the updated  
 DBNs. In order to avoid overloading the display of the results,  
 only the consumption in office 2, HQ\_office#2, is analyzed here.

Because of the density of the measurements, it is difficult to dis-  
 tinguish the differences between the fault situations and the base-  
 line, referenced as “E0.” Therefore, we limit the comparison of  
 the results to the period from 900 to 1800 h.

Figures 13 and 14 illustrate two examples of simulation of  
 energy consumption deviations in office 2 following the appear-  
 ance of faults. These figures show the deviations of the energy  
 consumption between the baseline E0 and two fault cases. The  
 first one (referenced as “case\_1”) consists in an increase in the  
 preset temperature of +1 K (see Fig. 13) and the second one  
 (references as “case\_2”) is a drift of the glazing transmission coef-  
 ficient of  $-0.5 \text{ W/m}^2 \text{ K}$  (see Fig. 14). In case\_1, the injected fault  
 on IT leads to an average relative discrepancy of performance  
 equal to  $+7.64\%$  (with a standard deviation of  $+6.34\%$ ). In case\_2  
 (injected fault on  $U_{wi}$ ), we observe respectively values of  
 $+1.86\%$  and  $+2.02\%$  for the average discrepancy and the stand-  
 ard deviation. One will note that the first fault has a higher effect  
 than the second one.

Figure 15 illustrates the probability of a process being under SC  
 $P(SC/x)$  over time in the case of an increase in preset temperature.  
 The faults are injected independently at time  $t_1 = 1000 \text{ h}$  and the  
 simulations are stopped at  $t_2 = 1800 \text{ h}$ . Here, it is worth pointing  
 out, since only one variable—HQ\_office#2—is analyzed, that the  
 $T^2$  Hotelling chart is reduced to one chart, “Xbar and S.” Given  
 that case\_1 is a simulated fault situation, a very rapid and very  
 clear drop of probability  $P(SC/x)$  can be seen after the injection of  
 faults at  $t_1 = 1000 \text{ h}$ . It is therefore very quick and easy to detect  
 fault situations. But, what is really of interest, and this is the  
 advantage of using BNs, is the ability to carry out a diagnosis, i.e.,  
 to determine the most probable initial cause of the fault.

For the same detection threshold (here, for example,  
 $P(SC/x) = 0.6$ ) and measurements obtained over a control period  
 of 6 h (c.f. Sec. 5.4. Influence of control period below for a

**Table 4 Simulation variables for fault situations**

Factors	Code	Low-performance threshold	High-performance threshold
AL (coefficient) <sup>a</sup>	AL	2	0.5
Indoor temperature or preset temperature (K) <sup>b</sup>	IT	+1	-1
Internal heat gains (coefficient) <sup>c</sup>	IG	1	2
Level of occupancy (coefficient) <sup>d</sup>	Pre	1	2
Albedo ( ) <sup>e</sup>	Alb	0.1	0.4
Glazing solar factor (coefficient) <sup>f</sup>	GF	-0.2	+0.2
Ground temperature (K)	GT	-3	+3
Low ceiling (coefficient) <sup>g</sup>	LC	0.5	1.2
Thermal bridges (coefficient) <sup>h</sup>	TB	1.2	0.8
$U_{\text{roof}}$ or High ceiling (coefficient) <sup>g</sup>	UR	0.5	1.2
$U_{\text{wall}}$ (coefficient) <sup>i</sup>	Uwa	0.78	1.62
$U_{\text{window}}$ (W/m <sup>2</sup> K) <sup>j</sup>	Uwi	3.2	4.2
Ventilation (kg/h) <sup>k</sup>	Ve	200	100

<sup>a</sup>AL: The air permeability coefficient of the nominal situation was taken to be equal to 1.7 m<sup>3</sup>/h.m<sup>2</sup>. The dispersion typically observed on buildings erected between 1948 and 2000 was retained, which corresponds to a disturbance of the default value by a multiplicative coefficient of between 0.5 and 2. For instance, the low performance level corresponds here to an air permeability of 3.4 m<sup>3</sup>/h.m<sup>2</sup>.

<sup>b</sup>Indoor or preset temperature: the preset temperature uncertainty equates to the measurement uncertainty of the indoor temperature, which covers sensor uncertainty and spatial sampling uncertainty, i.e., ±1 °C.

<sup>c</sup>Internal heat gains: it corresponds to heat power from electrical devices use. Its uncertainty can be large if not precisely measured.

<sup>d</sup>Level of occupancy: The occupancy measurement procedure (via motion detection) was affected by a bias consistently tending toward underestimation, due to the possible absence of motion during occupancy as well as the possible presence of multiple individuals in a single office. A disturbance of hourly occupancy of a factor of between 1 and 2 was therefore applied.

<sup>e</sup>Albedo: it is a coefficient (between 0 and 1) that corresponds to the reflective power of external surfaces around the building. The dispersion comes from the uncertainty of the value (that is usual not measured) and the diversity of surfaces around the building.

<sup>f</sup>Characteristics of the bays: to represent the uncertainty of the thermal and radiation characteristics of the bays, the modeling variables of the glazing system (woodwork included) are disturbed by the method described in the appendix. The transmission (total and visible) multiplication coefficient for the glazing was considered to be between 0.8 and 1.2: this corresponds to a solar disturbance factor of ±20% around the basic value of 0.81.

<sup>g</sup>Low and high ceilings: the corrective coefficients applied correspond to an insulation thickness of between 10 cm (0.5 × 20 cm) and 24 cm (1.2 × 20 cm); 20 cm being the nominal thickness.

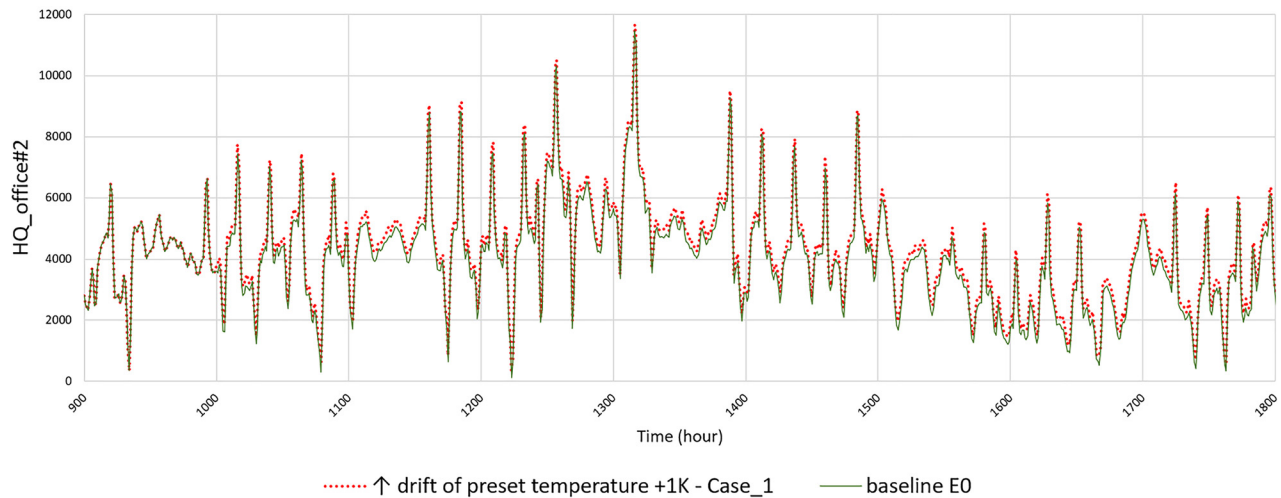
<sup>h</sup>Thermal bridges: the chosen dispersion deals with uncertainty of the materials and its implementation in the building.

<sup>i</sup>Exterior walls: the maximum dispersion of the thermal resistance of the air knives from the ThU rules was retained, i.e., between 0.11 and 0.23 m<sup>2</sup> K/W (nominal value  $U_{\text{wall}} = 7 \text{ W/m}^2 \text{ K}$ ).

<sup>j</sup>Windows: The amplitude of the disturbances applied to the conductivity of the filling gas and the thermal transmission coefficient of the woodwork resulted in a variation of transmission coefficient for the glazing system ( $U_{\text{windows}}$ ) of between 3.2 and 4.2 W/m<sup>2</sup> K.

<sup>k</sup>Nominal ventilation flow rate: ventilation flow rate was measured by the cone method. The measured value was 150 m<sup>3</sup>/h. The experimental measurement conditions led us to increase this uncertainty and retain a dispersion of between 100 and 200 m<sup>3</sup>/h.

**Case\_1: Simulated effect of a drift of the preset temperature**



**Fig. 13 Simulation of consumption deviations for an increase in preset temperature (case\_1)**

779 definition), the fault is more rapidly identified in fault case\_1  
 780 (21 h) than in fault case\_2 (463 h ≈ 19 days). This fact is intrinsi-  
 781 cally linked to the amplitude of the faults effects. In case-1 we  
 782 have a simulated average deviation of 249.7 and standard devia-  
 783 tion of 53.5, and in case\_2, a mean and standard deviation of 56.1  
 784 and 34.1, respectively.

Detection time is an important performance indicator. It 785  
 depends on various factors; of course, on the amplitude of the 786  
 fault effect on heat or energy needs HQ, but also on the value of 787  
 the detection threshold  $P(SCLx)$ , and on the width of the control 788  
 period. Detection time is also stochastic in that it depends on the 789  
 fault injection time. Figure 16 presents the distribution laws of 790



Case\_2: Simulated effect of a drift of the glazing transmission coefficient

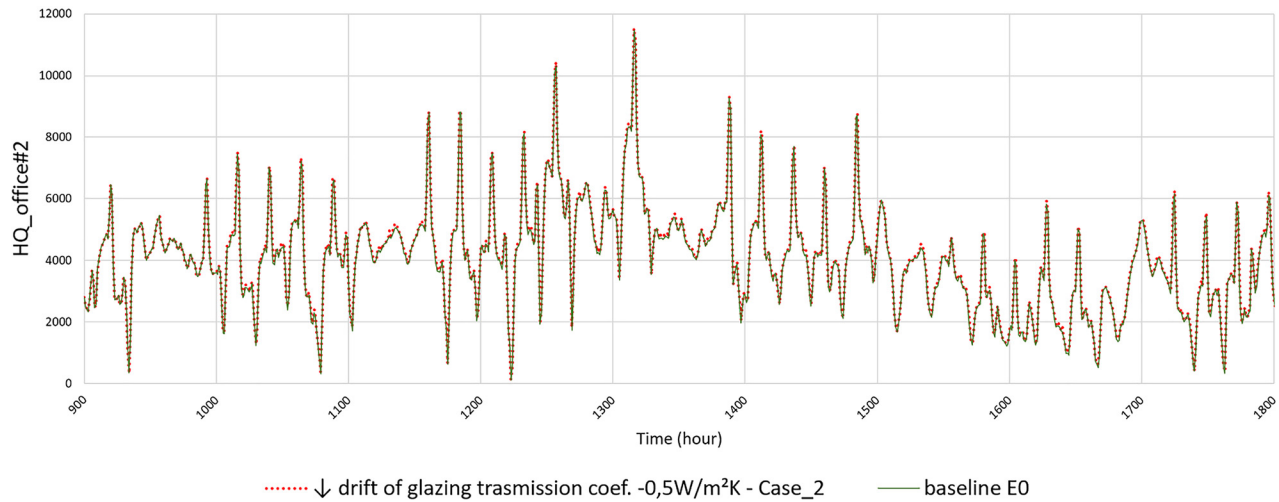


Fig. 14 Simulation of consumption deviations for a deterioration in glazing transmission coefficient (case\_2)

Simulations of individual drift effects (both Cases 1 & 2)

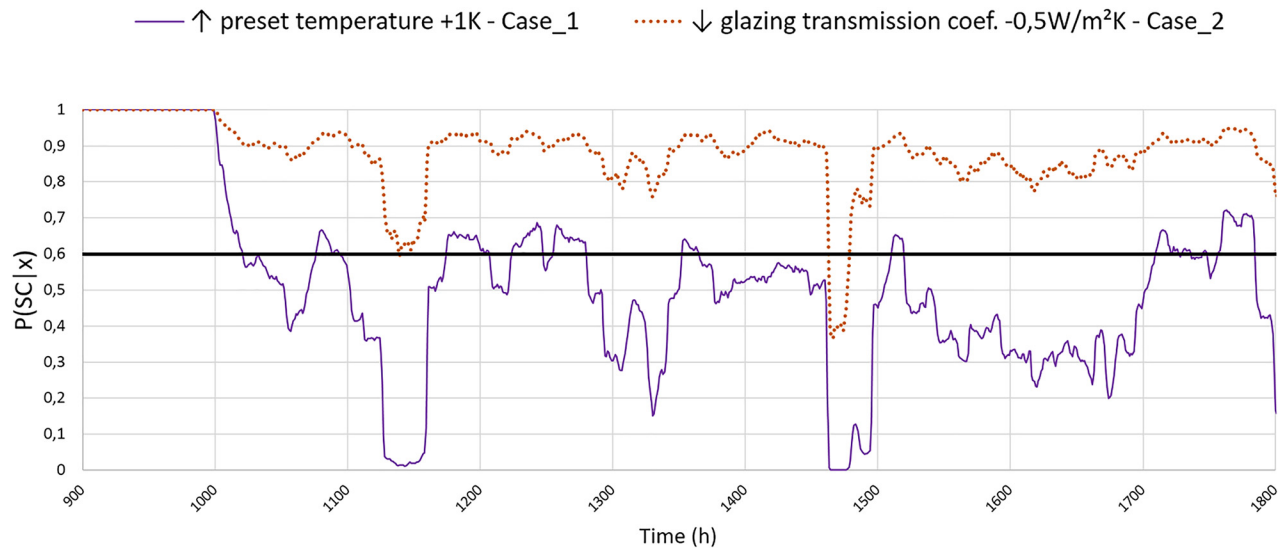


Fig. 15 Probability of being under SC after an increase in the preset temperature (case\_1) or a decrease in the glazing transmission coefficient (case\_2)

791 detection times for a preset temperature and a glazing transmission  
 792 coefficient drifts for statistically injected fault times taken,  
 793 for example, between 1000 and 2000 h and for the same detection  
 794 threshold of 0.6 and with a control period of 6 h. For a very signif-  
 795 icant effect on energy needs (case\_1 of a positive drift of the pre-  
 796 set temperature), detection times are short and their distribution  
 797 follows an exponential law. For a less significant effect (case\_2 of  
 798 negative drift of glazing transmission coefficient), detection times  
 799 are longer and their distribution follows a lognormal law.

800 If the detection time in case\_1 is shorter, it is more difficult to  
 801 identify the possible causes (and hierarchize them). In fact, the  
 802 more quickly the fault is identified, the more limited the quantity  
 803 of available information is, and the more random the characteriza-  
 804 tion of the fault will be. Table 5 shows a sorted list, obtained via a  
 805 BN coupled with control chart, of probable causes of faults. The  
 806 results have been standardized, and the sum of all the percentages  
 807 for all 13 variables should equal 100%.

808 A lack of discrimination of possible causes is plainly evident in  
 809 case\_1: while the first six causes listed do represent 82.8% of the

possibilities, the cause at the top of the list is only 2.3 times more  
 likely than the sixth one. Moreover, the IT variable which is the  
 actual cause of the fault [IT for indoor temperature (or preset)—  
 see Table 4] is not identified as being the most possible cause. In  
 this specific case\_1, we think that the weakness in characterizing  
 the fault is due to the fact that the deviations generated by the sim-  
 ulation in the first times after  $t_1$  were greater than the average of  
 249.7 (which is what was measured between the time the fault  
 was generated at 1000 h and the end of the simulation at 3400 h).

819 In case\_2, the drift has been characterized correctly. The detec-  
 820 tion threshold is still 0.6, and the drift is identified after 463 h of  
 821 operation in degraded mode: having recourse to more comprehen-  
 822 sive information, the BN offers clear identification of the cause of  
 823 drift, which is  $U_{wi}$  ( $U_{window}$  coefficient).  $U_{wa}$  is identified as the  
 824 second most possible cause, however, with 2.8 times less likeli-  
 825 hood of being the cause. It seems logical to assume that  $U_{wa}$   
 826 appears as a probable cause because the effect of injecting the  
 827 fault into  $U_{wa}$  only (in a simulation not shown here) gave a simu-  
 828 lated mean deviation of 61.8 and standard deviation of 39.8, in

### Distribution of detection times after fault injection (detection threshold $P(SC|x)=0.6$ and control period $\delta_c=6$ h)

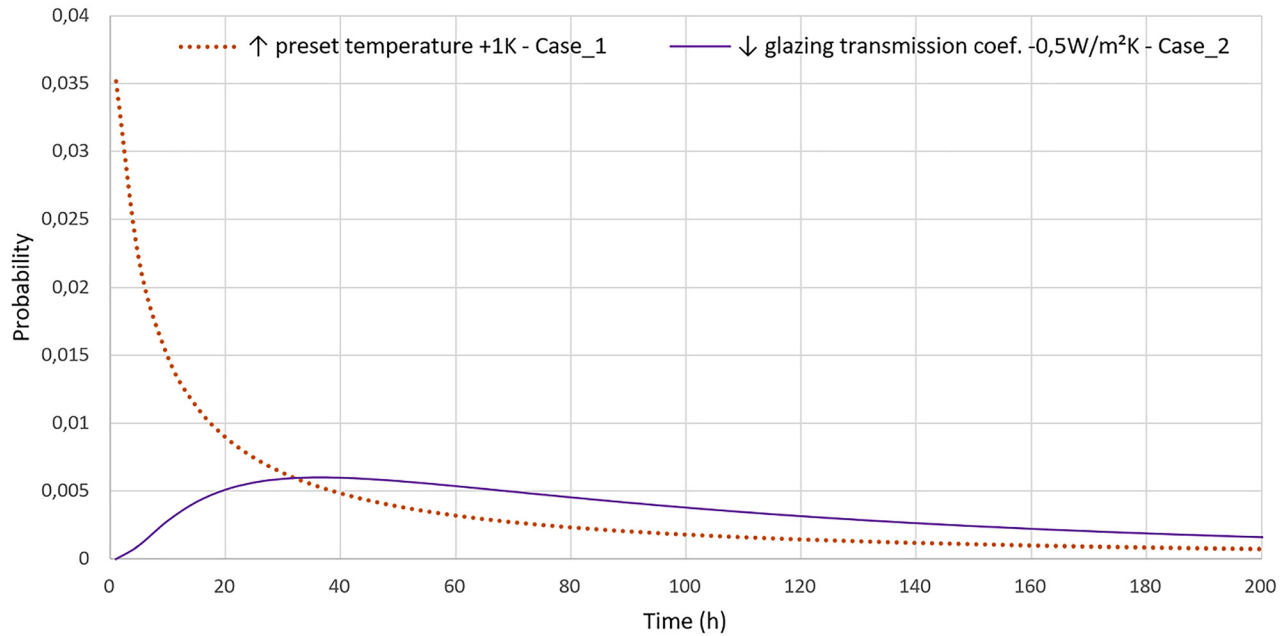


Fig. 16 Statistical distributions of detection times after an increase in the preset temperature (case\_1) or a decrease in the glazing transmission coefficient (case\_2)

Table 5 Ranking in descending order of possible causes of faults or drifts [detection threshold = 0.6 – control period = 6 h]

Fault/drift	Detection time (h)	Probable cause #1	Probable cause #2	Probable cause #3	Probable cause #4	Probable cause #5	Probable cause #6
IT (K) (or preset)	21	IG → 20.2%	IT → 17.0	Uwa → 13.3	Uwi → 13.0	TB → 10.7	AL → 8.6
$U_w$ (W/m <sup>2</sup> .K)	463	Uwi → 48.2%	Uwa → 17.1	TB → 6.5	AL → 4.0	UR → 2.9	Alb → 2.2

where AL is the air leakage, IT is the indoor or preset temperature, IG is the internal heat gains, Pre is the level of occupancy, Alb is the albedo, GF is the glazing solar factor, GT is the ground temperature, LC is the low ceiling, TB is the thermal bridges, UR is  $U_{roof}$  or high ceiling, Uwa is  $U_{wall}$ , Uwi is  $U_{window}$ , and Ve is the ventilation.

### Simulations of preset temperature fault effects (from +0.5K to +2K)

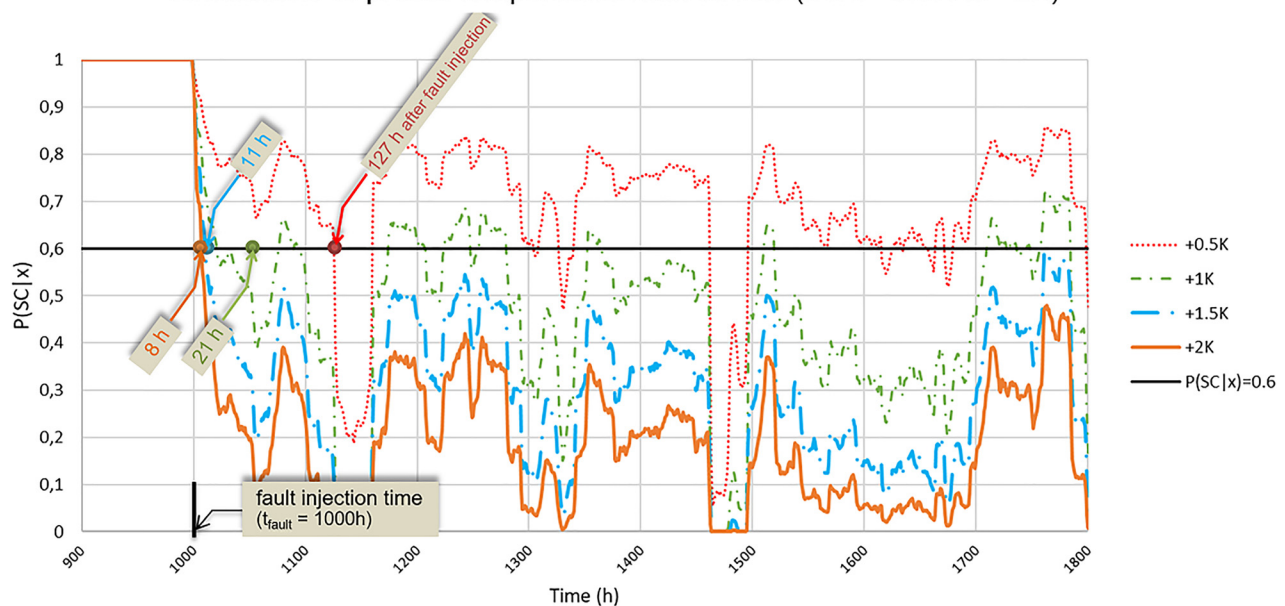


Fig. 17 Influence of the preset temperature drift amplitude on the probability of process being under SC

**Table 6 Influence of drift amplitude (case\_1 of preset temperature) on ranking of probable causes of faults [detection threshold = 0.6; control period = 6 h]**

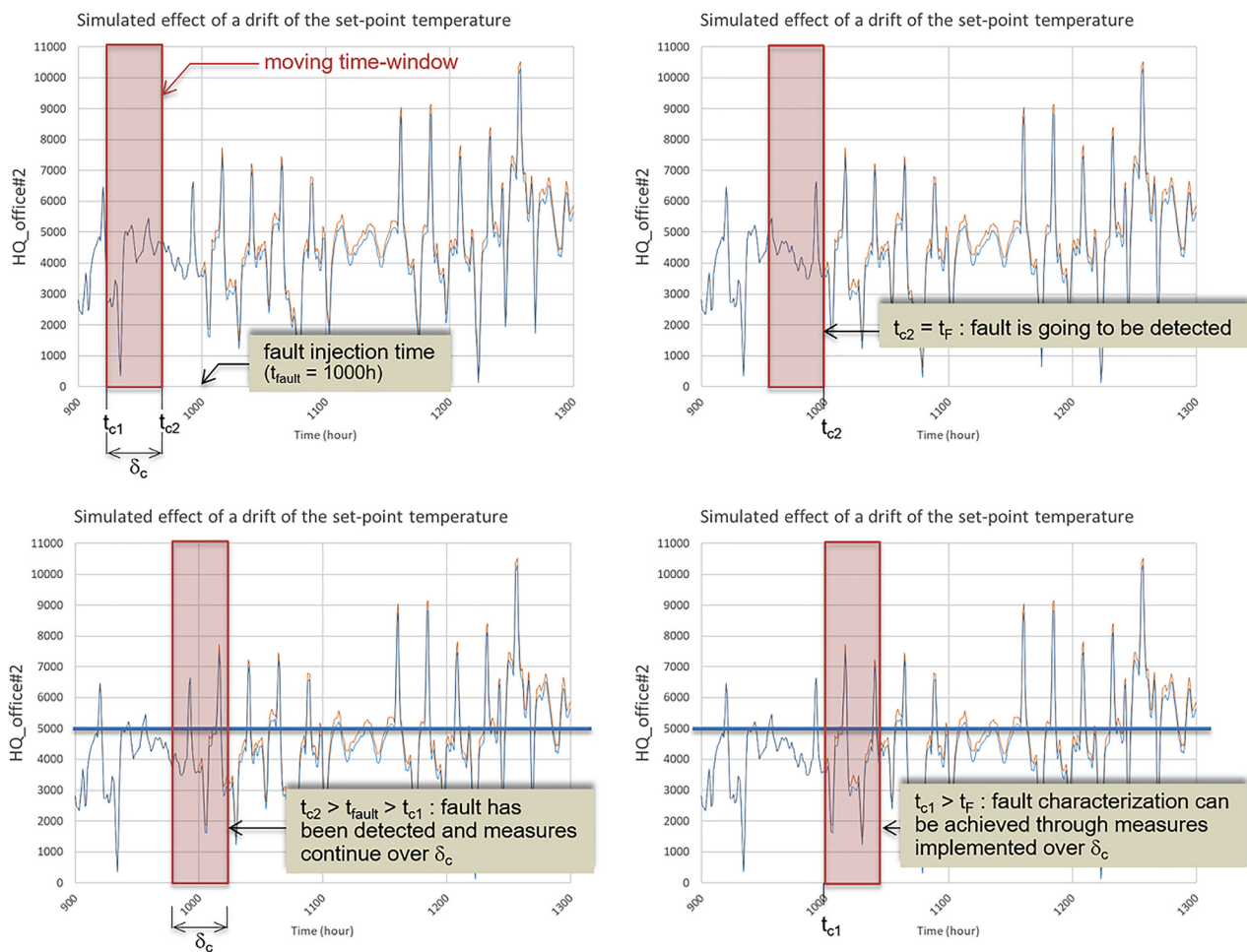
Drift amplitude	Detection time (h)	Probable cause #1	Probable cause #2	Probable cause #3	Probable cause #4	Probable cause #5	Probable cause #6
Drift of +0.5 K	127	IG → 19.1%	IT → 14.7%	Uwa → 12.2%	Uwi → 12.0%	TB → 11.2%	AL → 10.9%
Drift of +1 K	21	IG → 20.2%	IT → 17.0%	Uwa → 13.3%	Uwi → 13.0%	TB → 10.7%	AL → 8.6%
Drift of +1.5 K	11	IT → 38.1%	IG → 15.1%	Uwi → 11.2%	Uwa → 11.0%	TB → 6.2%	AL → 3.9%
Drift of +2 K	8	IT → 78.2%	IG → 5.1%	Uwi → 4.2%	Uwa → 3.0%	TB → 1.1%	AL → 0.8%

where AL is the air leakage, IT is the indoor or preset temperature, IG is the internal heat gains, Pre is the level of occupancy, Alb is the albedo, GF is the glazing solar factor, GT is the ground temperature, LC is the low ceiling, TB is the thermal bridges, UR is  $U_{roof}$  or high ceiling, Uwa is  $U_{wall}$ , Uwi is  $U_{window}$ , and Ve is the ventilation.

**Table 7 Influence of threshold value on ranking of probable causes of faults [control period of 6 h]**

Threshold value	Detection time (h)	Probable cause #1	Probable cause #2	Probable cause #3	Probable cause #4	Probable cause #5	Probable cause #6
Threshold of 0.6	21	IG → 20.2%	IT → 17.0%	Uwa → 13.3%	Uwi → 13.0%	TB → 10.7%	AL → 8.6%
Threshold of 0.5	53	IT → 40.1%	IG → 14.0%	Uwa → 12.2%	Uwi → 11.8%	TB → 8.8%	AL → 5.6%
Threshold of 0.4	114	IT → 78.2%	IG → 5.1%	Uwi → 4.2%	Uwa → 3.0%	TB → 1.1%	AL → 0.8%

where AL is the air leakage, IT is the indoor or preset temperature, IG is the internal heat gains, Pre is the level of occupancy, Alb is the albedo, GF is the glazing solar factor, GT is the ground temperature, LC is the low ceiling, TB is the thermal bridges, UR is  $U_{roof}$  or high ceiling, Uwa is  $U_{wall}$ , Uwi is  $U_{window}$ , and Ve is the ventilation.



**Fig. 18 Principle of the fault characterization moving time-window**

829 other words, values that are very close to the simulated ones for  
830  $U_{wi}$  only: 56.1 and 34.1.

831 **5.2 Influence of Drift Amplitude.** The influence of the drift  
832 amplitude has been suggested in the previous paragraphs. Here,

we study this influence by considering the effect of different 833  
values of positive preset temperature drifts. To some extent, this 834  
parallels to a steady decline of preset temperature control. Figure 835  
17 shows the probability  $P(SC|x)$  that the process is still under 836  
statistical control over time for different drift values from +0.5K to 837

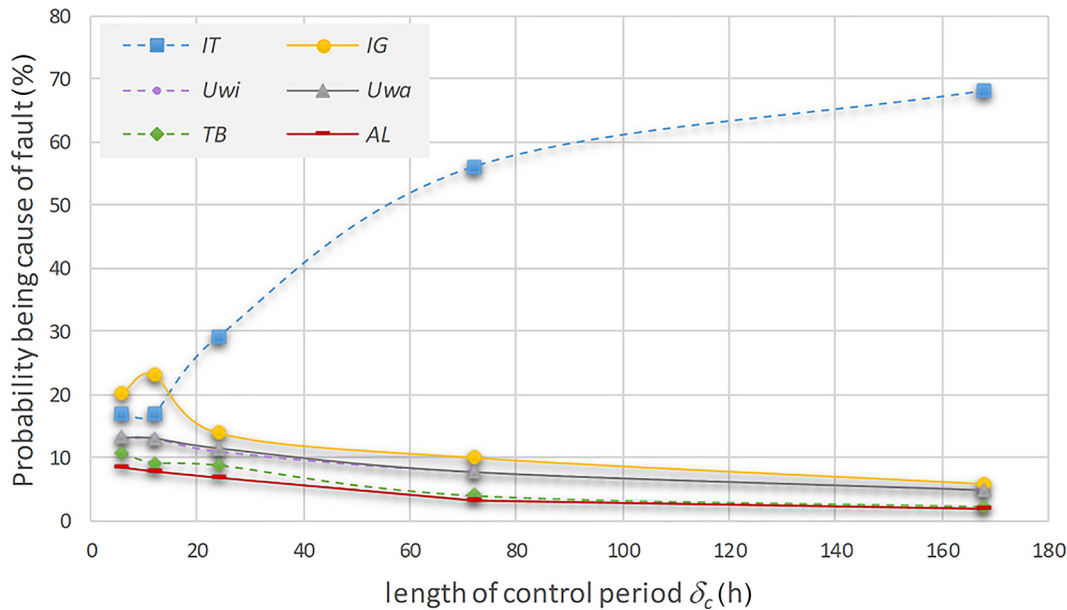


**Table 8 Influence of control period on the ranking of possible causes of faults [threshold of 0.6]**

Control period $\delta_c$	Probable cause #1	Probable cause #2	Probable cause #3	Probable cause #4	Probable cause #5	Probable cause #6
$\delta_c = 6$ h	IG $\rightarrow$ 20.2 %	IT $\rightarrow$ 17.0	Uwa $\rightarrow$ 13.3	Uwi $\rightarrow$ 13.0	TB $\rightarrow$ 10.7	AL $\rightarrow$ 8.6
$\delta_c = 12$ h	IG $\rightarrow$ 23.1 %	IT $\rightarrow$ 16.9	Uwi $\rightarrow$ 13.2	Uwa $\rightarrow$ 13.1	TB $\rightarrow$ 9.2	AL $\rightarrow$ 7.7
$\delta_c = 24$ h	IT $\rightarrow$ 29.1 %	IG $\rightarrow$ 14.0	Uwa $\rightarrow$ 11.6	Uwi $\rightarrow$ 11.1	TB $\rightarrow$ 8.8	AL $\rightarrow$ 6.8
$\delta_c = 72$ h	IT $\rightarrow$ 56.1 %	IG $\rightarrow$ 10.1	Uwi $\rightarrow$ 7.8	Uwa $\rightarrow$ 5.2	TB $\rightarrow$ 4.1	AL $\rightarrow$ 3.3
$\delta_c = 168$ h	IT $\rightarrow$ 68.2 %	IG $\rightarrow$ 6.0	Uwi $\rightarrow$ 4.9	Uwa $\rightarrow$ 4.0	TB $\rightarrow$ 2.4	AL $\rightarrow$ 1.8

where AL is the air leakage, IT is the indoor or preset temperature, IG is the internal heat gains, Pre is the level of occupancy, Alb is the albedo, GF is the glazing solar factor, GT is the ground temperature, LC is the low ceiling, TB is the thermal bridges, UR is  $U_{roof}$  or high ceiling, Uwa is  $U_{wall}$ , Uwi is  $U_{window}$ , and Ve is the ventilation.

**Influence of control period on the ranking of possible causes of faults [threshold of 0.6]**



**Fig. 19 Simulations of consumption deviations for the case with two simultaneous faults (IT and Uwi)**

**Table 9 Influence of threshold and control period on ranking of possible causes of faults for two faults appearing simultaneously (IT and Uwi)**

		Control threshold	
		0.75	0.6
Control period	$\delta_c = 48$ h	IT $\rightarrow$ 32.2%; Uwa $\rightarrow$ 14.3%; IG $\rightarrow$ 12.7% Uwi $\rightarrow$ 11.4%; TB $\rightarrow$ 7.1%; AL $\rightarrow$ 5.3%	IT $\rightarrow$ 40.0; Uwa $\rightarrow$ 11.4; Uwi $\rightarrow$ 11.3 IG $\rightarrow$ 9.8; TB $\rightarrow$ 5.8; AL $\rightarrow$ 3.5
	$\delta_c = 144$ h	IT $\rightarrow$ 37.2; Uwa $\rightarrow$ 12.6; Uwi $\rightarrow$ 12.4 IG $\rightarrow$ 10.7; TB $\rightarrow$ 6.4; AL $\rightarrow$ 4.0	IT $\rightarrow$ 43.6; Uwa $\rightarrow$ 10.0; Uwi $\rightarrow$ 9.4 IG $\rightarrow$ 8.7; TB $\rightarrow$ 5.0; AL $\rightarrow$ 3.1

where AL is the air leakage, IT is the indoor or preset temperature, IG is the internal heat gains, Pre is the level of occupancy, Alb is the albedo, GF is the glazing solar factor, GT is the ground temperature, LC is the low ceiling, TB is the thermal bridges, UR is  $U_{roof}$  or high ceiling, Uwa is  $U_{wall}$ , Uwi is  $U_{window}$ , and Ve is the ventilation.

838 +2 K with a step of 0.5 K. We still consider that, as an example,  
839 the detection threshold is 0.6 and the control period is 6 h. We  
840 also indicate the different detection times after fault injection  
841 (here, fault is arbitrarily injected at 1000 h).

842 Of course, detection time decreases when the value of drift  
843 increases but, as suggested in the previous paragraphs, there is a  
844 negative impact on the ability to identify the possible cause lead-  
845 ing to the detected fault. The results of simulations presented in  
846 Table 6 show this negative impact that will push to focus on the

necessary optimization of the control parameters (detection 847  
threshold, control period ...). 848

**5.3 Influence of Detection Threshold of Control Chart. 849**  
The results presented previously (see primarily comments on 850  
case\_1) suggest that the detection time has an impact on the qual- 851  
ity of fault characterization. A short detection time mechanically 852  
reduces the quantity of information available to the BN coupled 853

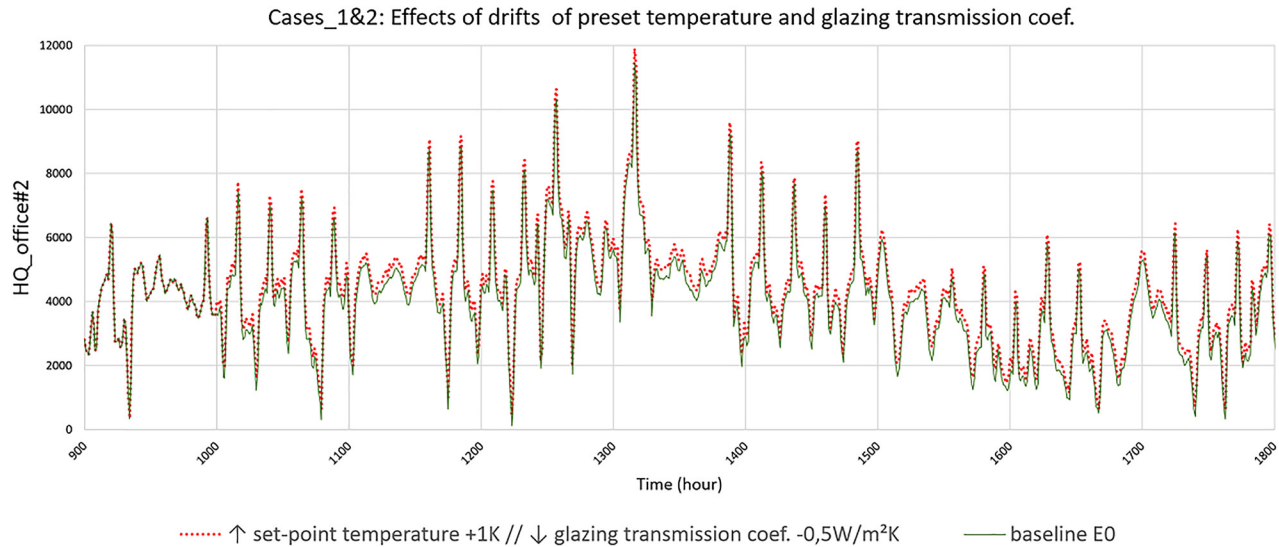


Fig. 20 Simulations of consumption deviations for the case with two simultaneous faults (IT and  $Uw$ )

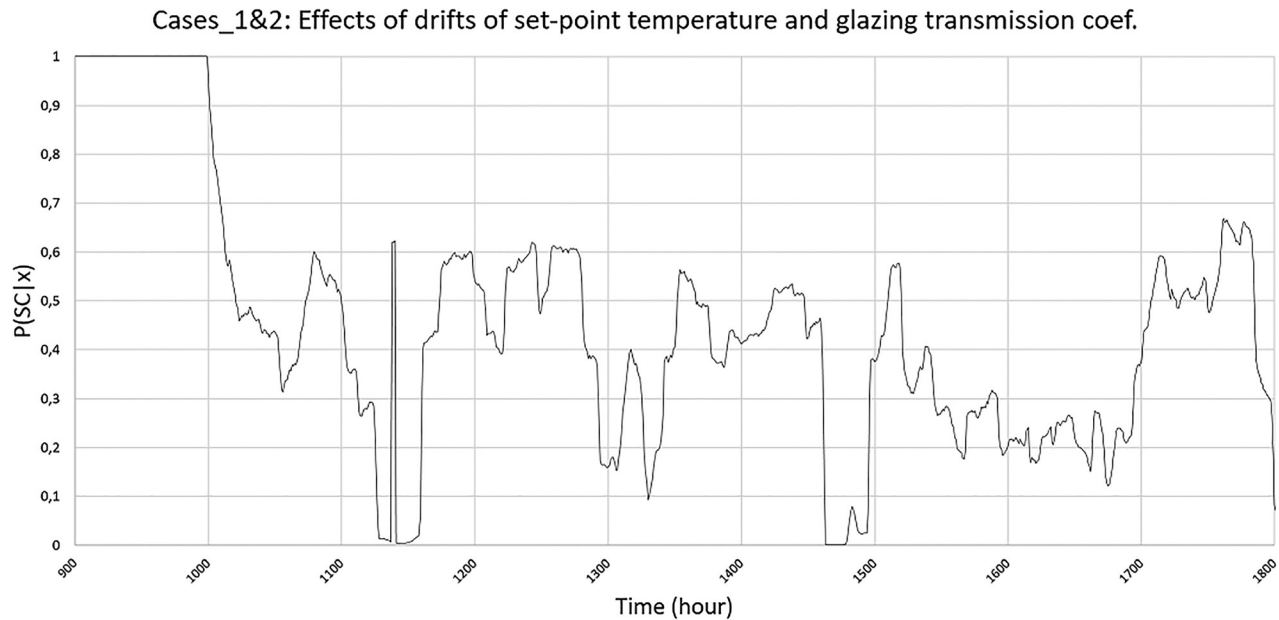


Fig. 21 Probability of process being under SC for a case with two faults appearing simultaneously (IT and  $Uw$ )

854 with control chart, thus increasing the ultimate risk of poor char- 872  
 855 acterization of the fault or drift. This detection time, which we 873  
 856 cannot a priori fully control, is linked to the detection threshold of 874  
 857 the control chart, which, however, can be controlled/chosen. In 875  
 858 Table 7 can be seen the influence of the chosen control chart 876  
 859 detection threshold on the characterization of the fault for the first 877  
 860 case\_1, dealt with previously. The threshold values of 0.6, 0.5, 878  
 861 and 0.4 were chosen arbitrarily to illustrate this influence. 879

862 We can see that the injected IT fault is identified at the 0.5 880  
 863 threshold (with a corresponding detection time of around 53 h) as 881  
 864 the most likely one (and 2.8 times more likely than the second 882  
 865 fault, IG). For a threshold value of 0.4 (with a corresponding 883  
 866 detection time of roughly 114 h), the first probable cause is unam- 884  
 867 biguously identified; the probability of the IT fault occurring 885  
 868 effectively overtakes that of IG with a ratio of over 15. This high 886  
 869 level of discrimination is because the impact of the IT fault alone 887  
 870 on consumption deviations is intrinsically far greater than the 888  
 871 effects of the other variables.

872 Finally, we must point out that in the example given in our 873  
 874 study the characterization error for low threshold values grows in 875  
 876 proportion to the degree to which the first simulated values differ 877  
 878 from the mean value calculated between times  $t_1$  and  $t_2$ , i.e., the 879  
 880 start and end times of the simulation. 881

882 **5.4 Influence of the Control Period.** The previous simula- 883  
 884 tions have demonstrated that modification of the threshold value 885  
 886 can improve the characterization of a fault. The aim of adjusting 887  
 888 the threshold value in this way is to mechanically increase the 889  
 890 duration over which consumption deviations are measured, thus 891  
 892 ensuring that comprehensive information is available and used. 893  
 894 Another strategy is possible, namely the adjustment of the control 895  
 896 period. The idea is that once a fault has been detected, the mea- 897  
 898 surement of deviations continues for a certain time period  $\delta_c$  899  
 899 in order to obtain a better fault “signature.” In practical terms, that 900  
 900 means to increase the time duration of the signal analysis, i.e., to 901  
 901 insert a time window into the signal (see Fig. 18). 902

889 For the threshold of 0.6, which evidently caused problems for  
 890 the characterization of the fault in simulation case\_1, we tested the  
 891 influence of the control period, i.e., duration  $\delta_c$ , on the identification  
 892 of probable causes. Table 8 sums up the results obtained for five val-  
 893 ues of control period. It shows the impact of the control period on  
 894 the rank and scale of the probabilities achieved when identifying  
 895 fault causes. The more the length of the control period  $\delta_c$  increases,  
 896 the better the actual cause of fault is detected. In Fig. 19, we show  
 897 the variation of the probability of being the cause of fault (for the  
 898 sixth first most probable causes) as a function of  $\delta_c$ . Factor IT is con-  
 899 sidered to be the most probable cause of fault as soon as  $\delta_c$  reaches  
 900 about 16 h. After 60 h of control period, the probability of factor IT  
 901 to be the cause of fault is five times higher than the one of the second  
 902 ranked possible cause of fault. This ratio exceed ten for  $\delta_c$  beyond  
 903 140 h. We might finally be inclined to choose a large control period  
 904 to be certain to detect the actual cause, but if this control period is  
 905 too long the consequence of the ongoing deterioration of perform-  
 906 ance might be too critical. The choice of the control period length is  
 907 always a matter of compromise (see Table 9).

908 The results of simulations in Table 9 show that the cause IT is  
 909 indeed identified as most likely regardless of the control combina-  
 910 tion (threshold/period). This is primarily due to the fact that fault  
 911 IT has a greater impact on consumption deviations than do the  
 912 other variables. As for cause *Uwi*, it does feature among the most  
 913 likely causes of drift but is never ranked in first or second place,  
 914 where it should be.

915 **5.5 Influence of Number of Faults.** We have demonstrated  
 916 the rich potential of BNs coupled to a control chart in the pres-  
 917 ence of a single fault—a potential conditional upon an appropriate  
 918 degree of adjustment of the thresholds and control period. Here,  
 919 we examine the appearance, after a period of 1000 h, of two faults  
 920 simultaneously, on variables IT and *Uwi*. The simulations that we  
 921 present for office 2 were carried out with two threshold values  
 922 (0.75 and 0.6) and two control periods (48 and 144 h). Figures 20  
 923 and 21 show the results obtained by simulation via BN. Table 9  
 924 synthesizes the ranking of probable causes of faults.

925 In this situation where two faults on IT and *Uwi* are simultane-  
 926 ously injected, we have found that the value of the average rela-  
 927 tive discrepancy reaches +9.06% with a standard deviation of  
 928 +7.41%. One remark that the two effects of the fault taken indi-  
 929 vidualy are almost added together (we recall that the average dis-  
 930 crepancies for case\_1 and case\_2 have been found, respectively,  
 931 at +7.64% and 1.86% (see Sec. 5.1)).

932 **6 Conclusion**

933 This paper proposes a hybrid FDD method for detecting faults  
 934 and energy performance drifts in buildings during its operation

and maintenance stage. The method is based on the graphical 935  
 method of BNs. It is referred as hybrid since we use actual data 936  
 for the BN construction in functional state and DES to create a 937  
 complementary database in dysfunctional states. We have shown 938  
 the potential of the BN in the process of detection and ranking of 939  
 the probable causes of a fault or drift of energy performance for 940  
 an actual building located in Les Ponts-de-Cé, France. 941

Our modeling approach is divided into three stages. In the first 942  
 one, we compile a whole-year database of hourly measured/actual 943  
 input variables related to environmental conditions, building 944  
 envelope and energy systems performances, level of occupancy 945  
 and outputs (energy needs) extracted from actual data and/or pre- 946  
 dicted from DES. The DES model of the building is validated by 947  
 comparisons with the measured energy needs. 948

In the second stage, we explore a number of alternative Bayes- 949  
 ian networks designed for the modeling of a building in opera- 950  
 tional mode. The BN is used to simulate a baseline of the energy 951  
 needs variations in this operational mode, as a function of input 952  
 variables. A robustness study allows us to reduce the size of the 953  
 database with small disturbances on the BN architecture and on 954  
 its efficiency. This database reduction and the consequent 955  
 decrease in the computation time facilitate the development of 956  
 DBN with continuous nodes for more accurate simulations and 957  
 better diagnosis and prognostic performances. This DBN is 958  
 updated in a third stage using DES simulating several types of 959  
 faults or drifts of the model inputs. Once the inference rules for 960  
 dysfunctional operating modes are constructed, we carry out FDD 961  
 simulations. We show the potential but also the limitations of 962  
 using this approach for the ranking of probable causes of an 963  
 energy performance fault. Some of these limitations can be 964  
 explained by imperfect optimization of the control period and the 965  
 threshold adjustments of the control charts. The suggested 966  
 approach seems to have a lower degree of accuracy when several 967  
 faults appear simultaneously. In our further works, we will try to 968  
 address these shortcomings. 969

**Acknowledgment** 970

This study was funded by the grants from the National Associa- 971  
 tion of Research (ANR) in France within the OMEGA project and 972  
 the data were provided by CEREMA (Centre d'Etudes et d'expertise 973  
 sur les Risques, l'Environnement, la Mobilité et 974  
 l'Aménagement). 975

**Funding Data** 976

- National Association of Research (10.13039/ 977  
 501100001665). 978

**Table 10 The 104 fault situations simulated to update the DBN [low, high, and nominal level of performance are respectively coded by numbers (-1), (+1) and (0)].**

Factors →	Air Leakage	Preset temperature	Internal heat Gains	Level of occupancy	Glazing Albedo	Glazing solar factor	Ground temperature	Low Ceiling	Thermal Bridges	$U_{\text{roof}}$ Or high ceiling	$U_{\text{wall}}$	$U_{\text{window}}$	Ventilation
Fault situation # ↓	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13
E1	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E2	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E3	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E4	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E5	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E6	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E7	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E8	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)
E9	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)
E10	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)
E11	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)
E12	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)



Table 10 (continued)

Factors →	Air Leakage	Preset temperature	Internal heat Gains	Level of occupancy	Glazing Albedo	Glazing solar factor	Ground temperature	Low Ceiling	Thermal Bridges	$U_{\text{roof}}$ Or high ceiling	$U_{\text{wall}}$	$U_{\text{window}}$	Ventilation
Fault situation # ↓	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13
E13	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)
E14	(+1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E15	(+1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E16	(+1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E17	(+1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E18	(+1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E19	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E20	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)
E21	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)
E22	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)
E23	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)
E24	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)
E25	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)
E26	0	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E27	(-1)	0	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E28	(-1)	(-1)	0	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E29	(-1)	(-1)	(-1)	0	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E30	(-1)	(-1)	(-1)	(-1)	0	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E31	(-1)	(-1)	(-1)	(-1)	(-1)	0	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E32	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	0	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E33	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	0	(-1)	(-1)	(-1)	(-1)	(-1)
E34	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	0	(-1)	(-1)	(-1)	(-1)
E35	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	0	(-1)	(-1)	(-1)
E36	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	0	(-1)	(-1)
E37	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	0	(-1)
E38	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	0
E39	(-1)	(+1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E40	(-1)	(+1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E41	(-1)	(+1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E42	(-1)	(+1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E43	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E44	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)
E45	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)
E46	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)
E47	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)
E48	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)
E49	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)
E50	(-1)	(-1)	(+1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E51	(-1)	(-1)	(+1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E52	(-1)	(-1)	(+1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E53	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E54	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)
E55	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)
E56	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)
E57	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)
E58	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)
E59	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)
E60	(-1)	(-1)	(-1)	(+1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E61	(-1)	(-1)	(-1)	(+1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E62	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E63	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)
E64	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)
E65	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)
E66	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)
E67	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)
E68	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)
E69	(-1)	(-1)	(-1)	(-1)	(+1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E70	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E71	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)
E72	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)
E73	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)
E74	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)
E75	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)
E76	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)
E77	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)
E78	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)
E79	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)
E80	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)

Table 10 (continued)

Factors →	Air Leakage	Preset temperature	Internal heat Gains	Level of occupancy	Glazing Albedo	Glazing solar factor	Ground temperature	Low Ceiling	Thermal Bridges	$U_{\text{roof}}$ Or high ceiling	$U_{\text{wall}}$	$U_{\text{window}}$	Ventilation
Fault situation # ↓	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13
E81	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)
E82	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)
E83	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)
E84	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)
E85	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)
E86	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)
E87	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)
E88	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)
E89	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)
E90	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(+1)	(-1)	(-1)	(-1)	(-1)
E91	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(+1)	(-1)	(-1)	(-1)
E92	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(+1)	(-1)	(-1)
E93	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(+1)	(-1)
E94	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(-1)	(+1)
E95	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(+1)	(-1)	(-1)	(-1)
E96	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(+1)	(-1)	(-1)
E97	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(+1)	(-1)
E98	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(-1)	(+1)
E99	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(+1)	(-1)	(-1)
E100	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(+1)	(-1)
E101	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(-1)	(+1)
E102	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(+1)	(-1)
E103	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(-1)	(+1)
E104	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(-1)	(+1)	(+1)

References

[1] de Wilde, P., 2014, "The Gap Between Predicted and Measured Energy Performance of Buildings: A Framework for Investigation," *Autom. Constr.*, **41**, pp. 40–49.

[2] van Dronkelaar, C., Dowson, M., Spataru, C., and Mumovic, D., 2016, "A Review of the Regulatory Energy Performance Gap and Its Underlying Causes in Non-Domestic Buildings," *Front. Mech. Eng.*, **1**(1), pp. 1–10.

[3] Titikpina, F., Caucheteux, A., Charki, A., and Bigaud, A., 2015, "Uncertainty Assessment in Building Energy Performance With a Simplified Model," *Int. J. Metrol. Qual. Eng.*, **6**(3), p. 308.

[4] Katipamula, S., and Brambley, M. R., 2005, "Review Article: Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems—Part I: A Review," *HVACR Res.*, **11**(1), pp. 3–25.

[5] Jing, R., Wang, M., Zhang, R., Li, N., and Zhao, Y., 2017, "A Study on Energy Performance of 30 Commercial Office Buildings in Hong Kong," *Energy Build.*, **144**, pp. 117–128.

[6] Bynum, J. D., Claridge, D. E., and Curtin, J. M., 2012, "Development and Testing of an Automated Building Commissioning Analysis Tool (ABCAT)," *Energy Build.*, **55**, pp. 607–617.

[7] Wang, L., 2012, "Modeling and Simulation of HVAC Faulty Operations and Performance Degradation Due to Maintenance Issues," Asia Conference of International Building Performance Simulation Association (ASIM'2012), Shanghai, China, pp. 8.

[8] Verhelst, J., van Ham, G., Saelens, D., and Hensen, L., 2017, "Model Selection for Continuous Commissioning of HVAC-Systems in Office Buildings: A Review," *Renewable Sustainable Energy Rev.*, **76**, pp. 673–686.

[9] Yu, Y. B., Woradachjumboon, D., and Yu, D. H., 2014, "A Review of Fault Detection and Diagnosis Methodologies on Air-Handling Units," *Energy Build.*, **82**, pp. 550–562.

[10] Zhao, Y., Wang, S. W., and Xiao, F., 2013, "A Statistical Fault Detection and Diagnosis Method for Centrifugal Chillers Based on Exponentially-Weighted Moving Average Control Charts and Support Vector Regression," *Appl. Therm. Eng.*, **51**(1–2), pp. 560–572.

[11] Verbert, K., Babuska, R., and De Schutter, B., 2017, "Combining Knowledge and Historical Data for System-Level Fault Diagnosis of HVAC Systems," *Eng. Appl. Artif. Intell.*, **59**, pp. 260–273.

[12] Dong, B., O'Neill, Z., and Li, Z., 2014, "A BIM-Enabled Information Infrastructure for Building Energy Fault Detection and Diagnostics," *Autom. Constr.*, **44**, pp. 197–211.

[13] Wall, J., and Guo, Y., 2018, "RP1026: Evaluation of Next-Generation Automated Fault Detection & Diagnosis (FDD) Tools for Commercial Building Energy Efficiency—Part I: FDD Case Studies in Australia," *Low Carbon Living*, CRC, pp. 66.

[14] Abdollahi, A., Pattipati, K. R., Kodali, A., Singh, S., Zhang, S., and Luh, P. B., 2016, "Probabilistic Graphical Models for Fault Diagnosis in Complex Systems," *Principles of Performance Reliability Modeling and Evaluation—Essays in Honor of Kishor Trivedi on His 70th Birthday* (Springer Series in Reliability Engineering), L. Fiondella and A. Puliafito eds., pp. 109–139.

[15] Hao, J., Kang, J., Li, J., and Zhao, Z., 2012, "A Physical Model Based Research for Fault Diagnosis of Gear Crack," International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering, Chengdu, China, pp. 572–575.

[16] Schein, J., Bushby, S. T., Castro, N. S., and House, J. M., 2006, "A Rule-Based Fault Detection Method for Air Handling Units," *Energy Build.*, **38**(12), pp. 1485–1492.

[17] Cai, B., Huang, L., and Xie, M., 2017, "Bayesian Networks in Fault Diagnosis," *IEEE Trans. Ind. Inf.*, **13**(5), pp. 2227–2240.

[18] Afram, A., Janabi-Sharifi, F., Fung, A. S., and Raahemifar, K., 2017, "Artificial Neural Network (ANN) Based Model Predictive Control (MPC) and Optimization of HVAC Systems: A State of the Art Review and Case Study of a Residential HVAC System," *Energy Build.*, **141**, pp. 96–113.

[19] Li, G., and Hu, Y., 2019, "An Enhanced PCA-Based Chiller Sensor Fault Detection Method Using Ensemble Empirical Mode Decomposition Based Denoising," *Energy Build.*, **183**, pp. 311–324.

[20] Beghi, A., Cecchinato, L., Corazzol, C., Rampazzo, M., Simmini, F., and Susto, G. A., 2014, "A One-Class SVM Based Tool for Machine Learning Novelty Detection in HVAC Chiller Systems," *IFAC Proc. Vol.*, **47**(3), pp. 1953–1958.

[21] van Every, P. M., Rodriguez, M., Jones, C. B., Mammoli, A. A., and Martinez-Ramón, M., 2017, "Advanced Detection of HVAC Faults Using Unsupervised SVM Novelty Detection and Gaussian Process Models," *Energy Build.*, **149**, pp. 216–224.

[22] He, S., Zhiwei, W., Zhanwei, W., Xiaowei, G., and Zengfeng, Y., 2016, "Fault Detection and Diagnosis of Chiller Using Bayesian Network Classifier With Probabilistic Boundary," *Appl. Therm. Eng.*, **107**(1), pp. 37–47.

[23] Du, Z., and Jin, X., 2008, "Multiple Faults Diagnosis for Sensors in Air Handling Unit Using Fisher Discriminant Analysis," *Energy Convers. Manage.*, **49**(12), pp. 3654–3665.

[24] Kim, W., and Katipamula, S., 2018, "A Review of Fault Detection and Diagnostics Methods for Building Systems," *Sci. Technol. Built Environ.*, **24**(1), pp. 3–21.

[25] Uusitalo, L., 2007, "Advantages and Challenges of Bayesian Networks in Environmental Modelling," *Ecol. Modell.*, **203**(3–4), pp. 312–318.

[26] Taal, A., Itard, L., and Zeiler, W., 2018, "A Reference Architecture for the Integration of Automated Energy Performance Fault Diagnosis Into HVAC Systems," *Energy Build.*, **179**(15), pp. 144–155.

[27] Wang, Z., Wang, Z., Gu, X., He, S., and Yan, Z., 2018, "Feature Selection Based on Bayesian Network for Chiller Fault Diagnosis From the Perspective of Field Applications," *Appl. Therm. Eng.*, **129**(25), pp. 674–683.

[28] Zhao, Y., Wen, J., and Wang, S.-W., 2015, "Diagnostic Bayesian Networks for Diagnosing Air Handling Units Faults—Part II: Faults in Coils and Sensors," *Appl. Therm. Eng.*, **90**(5), pp. 145–157.

[29] Zhao, Y., Wen, J., Xiao, F., Yang, X., and Wang, S.-W., 2017, "Diagnostic Bayesian Networks for Diagnosing Air Handling Units Faults—Part I: Faults in Dampers, Fans, Filters and Sensors," *Appl. Therm. Eng.*, **111**(25), pp. 1272–1286.

[30] Cai, B., Liu, Y., Fan, Q., Zhang, Y., Liu, Z., Yu, S., and Ji, R., 2014, "Multi-Source Information Fusion Based Fault Diagnosis of Ground-Source Heat Pump Using Bayesian Network," *Appl. Energy*, **114**, pp. 1–9.

1044 [31] Marvin, H. J., Bouzembrak, Y., Janssen, E. M., van der Zande, M., Murphy, F.,  
 1045 Sheehan, B., Mullins, M., and Bouwmeester, H., 2017, "Application of Bayesian  
 1046 Risk Assessment," *Nanotoxicology*, **11**(1), pp. 123–133.

1047 [32] Millán, E., Descalço, L., Castillo, G., Oliveira, P., and Diogo, S., 2013, "Using  
 1048 Bayesian Networks to Improve Knowledge Assessment," *Comput. Educ.*,  
**60**(1), pp. 436–447.

1049 [33] Shute, V., and Wang, L., 2016, "Assessing and Supporting Hard-to-Measure  
 1050 Constructs in Video Games," *The Wiley Handbook of Cognition and Assessment: Frameworks, Methodologies, and Applications*, ■, ■, pp. 535–562.

1051 [34] Lyons, D. M., Arkin, R. C., Jiang, S., O'Brien, M., Tang, F., and Tang, P.,  
 1052 2017, "Performance Verification for Robot Missions in Uncertain Environments,"  
*Rob. Auton. Syst.*, **98**, pp. 89–104.

1053 [35] Ju, Z., Ji, X., Li, J., and Liu, H., 2017, "An Integrative Framework of Human  
 1054 Hand Gesture Segmentation for Human-Robot Interaction," *IEEE Syst. J.*,  
**11**(3), pp. 1326–1336.

1055 [36] Slanzi, D., and Poli, I., 2014, "Evolutionary Bayesian Network Design for High  
 Dimensional Experiments," *Chemom. Intell. Lab. Syst.*, **135**(15), pp. 172–182.

1056 [37] Taylor, D., Biedermann, A., Hicks, T., and Champod, C., 2018, "A Template for  
 1057 Constructing Bayesian Networks in Forensic Biology Cases When Considering  
 Activity Level Propositions," *Forensic Sci. Int.: Genet.*, **3**, pp. 136–146.

1058 [38] Szkuta, B., Ballantyne, K. N., Kokshoorn, B., and van Oorschot, R. A. H., 2018,  
 1059 "Transfer and Persistence of Non-Self DNA on Hands Over Time: Using  
 1060 Empirical Data to Evaluate DNA Evidence Given Activity Level Propositions,"  
*Forensic Sci. Int.: Genet.*, **33**, pp. 84–97.

1061 [39] Bae, S.-C., and Lee, Y. H., 2018, "Comparative Efficacy and Tolerability of  
 1062 Monotherapy With Leflunomide or Tacrolimus for the Treatment of Rheumatoid  
 1063 Arthritis: A Bayesian Network Meta-Analysis of Randomized Controlled  
 Trials," *Clin. Rheumatol.*, **37**(2), pp. 323–330.

1064 [40] Chiremsel, Z., Nait Said, R., and Chiremsel, R., 2016, "Probabilistic Fault  
 1065 Diagnosis of Safety Instrumented Systems Based on Fault Tree Analysis and  
 Bayesian Network," *J. Failure Anal. Prev.*, **16**(5), pp. 747–760.

1066 [41] Sousa, H. S., Prieto-Castrillo, F., Matos, J. C., Branco, J. M., and Lourenço, P.  
 1067 B., 2018, "Combination of Expert Decision and Learned Based Bayesian Net-  
 1068 works for Multi-Scale Mechanical Analysis of Timber Elements," *Expert Syst. Appl.*,  
**93**(1), pp. 156–168.

AQ10 1069 [42] Bishop, C. M., 2006, *Pattern Recognition and Machine Learning*, Springer, ■,  
 p. 738.

1070 [43] Dondelinger, F., Lèbre, S., and Husmeier, D., 2013, "Non-Homogeneous  
 1071 Dynamic Bayesian Networks With Bayesian Regularization for Inferring Gene  
 1072 Regulatory Networks With Gradually Time-Varying Structure," *Mach. Learn.*,  
**90**(2), pp. 191–230.

1073 [44] Kwisthout, J., 2018, "Approximate Inference in Bayesian Networks: Parameter-  
 ized Complexity Results," *Int. J. Approximate Reasoning*, **93**, pp. 119–131.

1074 [45] Wu, P. P.-Y., Julian Caley, M., Kendrick, G. A., McMahon, K., and Mengersen,  
 1075 K., 2018, "Dynamic Bayesian Network Inferencing for Non-Homogeneous  
 Complex Systems," *J. R. Stat. Soc. Ser. C: Appl. Stat.*, **67**(2), pp. 417–434.

1076 [46] Black, A., Korb, K. B., and Nicholson, A. E., 2014, "Intrinsic Learning of  
 1077 Dynamic Bayesian Networks," *PRICAI 2014: Trends in Artificial Intelligence*  
 (Lecture Notes in Computer Science), Vol. 8862, ■, ■, pp. 256–269.

1078 [47] Hu, M., Chen, H., Shen, L., Li, G., Guo, Y., Li, H., Li, J., and Hu, W., 2018, "A  
 1079 Machine Learning Bayesian Network for Refrigerant Charge Faults of  
 1080 Variable Refrigerant Flow Air Conditioning System," *Energy Build.*, **158**, pp.  
 668–676.

1081 [48] Lin, S., Chen, X., and Wang, Q., 2018, "Fault Diagnosis Model Based on  
 1082 Bayesian Network Considering Information Uncertainty and Its Application in  
 1083 Traction Power Supply System," *IEEJ Trans. Electr. Electron. Eng.*, **13**(5), pp.  
 671–680.

1084 [49] Verrou, S., 2007, "Diagnosis and Monitoring of Complex Processes Via Bayesian  
 Networks," Ph.D. thesis, University of Angers, Angers, France.

1085 [50] Pillet, M., Boukar, A., Pairel, E., Rizzon, B., Boudaoud, N., and Cherfi, Z.,  
 1086 2013, "Multivariate SPC for Total Inertial Tolerancing," *Int. J. Metrol. Qual. Eng.*,  
**4**(3), pp. 169–175.

[51] Caucheteux, A., Sabar, A. E., and Boucher, V., 2013, "Occupancy Measure-  
 ment in Building: A Literature Review, Application on an Energy Efficiency  
 Research Demonstration Building," *Int. J. Metrol. Qual. Eng.*, **4**(2), pp.  
 135–144.

[52] ASHRAE Guideline 14—American Society of Heating, Refrigeration and Air  
 Conditioning Engineers, 2014, "ASHRAE Guideline 14 for Measurement of  
 Energy and Demand Savings," Atlanta, GA.

[53] EVO—Efficiency Valuation Organization, 2012, "International Performance  
 Measurement and Verification Protocol: Concepts and Options for Determining  
 Energy and Water Savings," ■, ■, Report No. EVO 10000 – 1:2012.

[54] WMO, 2008, *Guide to Meteorological Instruments and Methods of Observa-  
 tion*, 7th ed., World Meteorological Organization, Geneva, Switzerland.

[55] ASHRAE—American Society of Heating, Refrigeration and Air Conditioning  
 Engineers, 2017, "Thermal Environmental Conditions for Human Occupancy,"  
 ASHRAE, Atlanta, GA, Standard No. 55.

[56] Caucheteux, A., Gautier, A., and Lahrech, R., 2016, "A Metamodel-Based  
 Methodology for an Energy Savings Uncertainty Assessment of Building Retro-  
 fitting," *Int. J. Metrol. Qual. Eng.*, **7**(4), p. 402.

[57] Cali, D., Matthes, P., Huchtemann, K., Streblow, R., and Müller, D., 2015,  
 "CO2 Based Occupancy Detection Algorithm: Experimental Analysis and  
 Validation for Office and Residential Buildings," *Building Environ.*, **86**, pp.  
 39–49.

[58] Ansanay-Alex, G., Abdelouadoud, Y., and Schetelat, P., 2016, "Statistical and  
 Stochastic Modelling of French Households and Their Energy Consuming  
 Activities," 12th REHVA World Congress-CLIMA, ■, ■, ■.

[59] Yan, D., O'Brien, W., Hong, T., Feng, X., Gunay, H. B., Tahmasebi, F., and  
 Mahdavi, A., 2015, "Occupant Behavior Modeling for Building Performance  
 Simulation: Current State and Future Challenges," *Energy Build.*, **107**, pp.  
 264–278.

[60] Murphy, K., 2001, "The Bayesian Network Toolbox for Matlab," ■, ■, ■,  
<https://www.cs.ubc.ca/~murphyk/Papers/bnt.pdf>

[61] Lauritzen, S. L., 1992, "Propagation of Probabilities, Means and Variances in  
 Mixed Graphical Association Models," *J. Am. Stat. Assoc.*, **87**(420), pp.  
 1098–1108.

[62] Lauritzen, S. L., and Jensen, F., 2001, "Stable Local Computation With Condi-  
 tional Gaussian Distributions," *Stat. Comput.*, **11**(2), pp. 191–203.

[63] Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D. A., and Nolan, G. P., 2005,  
 "Causal Protein-Signaling Networks Derived From Multiparameter Single-Cell  
 Data," *Science*, **308**(5721), pp. 523–529.

[64] Claeskens, G., and Hjort, N. L., 2008, *Model Selection and Model Averaging*  
 (Part of Cambridge Series in Statistical and Probabilistic Mathematics), Cam-  
 bridge Press, Cambridge, UK, p. 332.

[65] Ghahramani, Z., 2001, "An Introduction to Hidden Markov Models and Bayes-  
 ian Networks," *Int. J. Pattern Recognit. Artif. Intell.*, **15**(1), pp. 9–42.

[66] Vlachopoulou, M., Chin, G., Fulle, J., and Lu, S., 2014, "Aggregated Residential  
 Load Modeling Using Dynamic Bayesian Networks," IEEE International  
 Conference on Smart Grid Communications, SmartGridComm, ■, ■, pp.  
 818–823.

[67] Ghahramani, A., Tang, C., Yang, Z., and Becerik-Gerber, B., 2015, "A Study of  
 Time-Dependent Variations in Personal Thermal Comfort Via a Dynamic  
 Bayesian Network," First International Symposium on Sustainable Human-  
 Building Ecosystems, Pittsburgh, PA, Oct. 5–6, 2015, pp. 99–107.

[68] Markovic, R., Wolf, S., Cao, J., Spinnräker, E., Wölki, D., Frisch, J., and van  
 Treeck, C., 2017, "Comparison of Different Classification Algorithms for the  
 Detection of User's Interaction With Windows in Office Buildings," Interna-  
 tional Conference on Future Buildings and Districts—Energy Efficiency From  
 Nano to Urban Scale (CISBAT), Lausanne, Switzerland, Sept. 6–8, pp.  
 337–342.

[69] Cai, B., Liu, Y., Ma, Y., Huang, L., and Liu, Z., 2015, "A Framework for the  
 Reliability Evaluation of Grid-Connected Photovoltaic Systems in the Presence  
 of Intermittent Faults," *Energy*, **93**, pp. 1308–1320.

[70] Hastie, T., Efron, B., 2012, "Lars: Least Angle Regression, Lasso and Forward  
 Stage Wise. R Package Version 1.1," ■, ■.