HAL open science

Use of Multidimensional Scaling for Fault Detection or Monitoring Support In A Continuous Commissioning

Hugo Geoffroy, Julien Berger, Benoît Colange, Sylvain Lespinats, Denys Dutykh, Gérard Sauce, Catherine Buhé

To cite this version:

Hugo Geoffroy, Julien Berger, Benoît Colange, Sylvain Lespinats, Denys Dutykh, et al.. Use of Multidimensional Scaling for Fault Detection or Monitoring Support In A Continuous Commissioning. Proceedings of the 16th IBPSA Conference, IBPSA, Sep 2019, Rome, Italy. pp.877-884, 10.26868/25222708.2019.210699 . hal-02411240

HAL Id: hal-02411240

https://hal.science/hal-02411240v1

Submitted on 20 Feb 2020

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.

Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

Use of Multidimensional Scaling for Fault Detection or Monitoring Support in a Continuous Commissionin

Building Simulation 2019 Conference

Hugo GEOFFROY[1], Julien BERGER[1], Benot COLANGE[2,3], Sylvain LESPINATS[2], Denys DUTYKH[3], Catherine BUHÉ[1], Gérard SAUCE[1]

[1] Univ. Grenoble Alpes, Univ. Savoie Mont-Blanc, CNRS, LOCIE, 73000 Chambry, France.

[2] Univ. Grenoble Alpes, INES, F-73375, Le Bourget du Lac, France.

[3] Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LAMA, 73000 Chambry, France.

Abstract

It is well known that the building energy use represents a significant part of the total energy use, *ca.* 40% in USA according to the Building Energy Data Book. With the improvement of new construction's efficiency, the share of equipment's energy use increases more and more compared to the overall building energy use. This paper proposes to study a new approach in building's systems Fault Detection and Diagnosis (FDD), so as to provide an intuitive FDD tool for every operator of the maintenance staff regardless of their qualifications. This new approach uses a method of multivariate statistics, providing easily understandable outputs allowing a quick comprehension of the equipment fault by building maintenance staff. Thereby the number of unsolved problems can be minimized and the intervention time would be considerably reduced, avoiding unexpected energy use and equipment's premature obsolescence.

Introduction

Nowadays, thanks to more and more ambitious regulations in terms of energy efficiency for both existing and new construction, we are required to achieve significant reduction of the energy use in the building sector. Nevertheless, the expected energy use of buildings remains under the estimations after comparison with in situ measurements (de Wilde, 2014). One key explanation for this problem is the malfunctioning of systems or their controls (Katipamula and Brambley, 2005). The reduction of these unwanted energy use can be reduced by Fault Detection and Diagnosis (FDD) (Li and O'Neill, 2018). It aims at detecting the occurrence of malfunctions during the equipment operation, to inform operation or maintenance staff in the shortest possible time to avoid violating indoors comfort or too high energy use. Since the creation of the "Annex 25" (Hyvarinen et al., 1999) of the International Energy Agency (IEA), the real implementation of this tools in buildings areas remains an open challenge. Furthermore, with the development of smart buildings, a new challenge appears (Lazarova-Molnar et al., 2016). There is a big amount of data to deal with and old methods are not always able to cope with it. Indeed, even in residential area, it may produce a huge amount of data for regulation of simplest systems, as Air Handling Unit (AHU) with numerous measured data recorded at short time step. This context increases the computing time and makes the FDD complex or too expensive for residential buildings.

The FDD techniques in the building sector have been widely described and classified during the last years (Bruton et al., 2014; Kim and Katipamula, 2018). In this study we chose to

focus on AHU and their Fault Detection and Diagnosis methods which can be divided in four main branches (Yu et al., 2014) : analytical-based methods, knowledge-based methods, data-driven methods, and a combination of all of those. The analytical-based methods usually compare measured data to the output of a mathematical model (*i.e.* first approach) and use the residual to detect faults in systems. The advantage of this method is that, unknown faults may be detected without huge quantity of measured data. However, the models which gives the required information can be complex and/or inaccurate. The knowledge-based methods use Artificial Intelligence (AI) along with expert analysis, to extract knowledge from data to detect a fault. These methods do not need model to detect and diagnose a fault, but have to be based on labeled historical data (to apply the AI) and usually did not detect unknown faults. The data-driven methods find relations among data patterns and identified faults. The key interest of these methods is that they do not require complex model or informations to detect fault, even unknown, but the capacity to diagnose can be reduced. Combination of these different methods appeared in recent years to solve the inherent problems of individual techniques.

Nevertheless, it is still an open challenge for the real time FDD tools to present the results in a satisfying and intelligible form. It is a major issue to propose a tool understandable to all maintenance operators regardless their qualification to analyze the FDD tools outputs (Djuric and Novakovic, 2009).

Aims

The aims of this publication are to explore a new Fault Detection and Diagnosis method for the building area and to demonstrate the potential of MultiDimensional Scaling (MDS) in a FDD tool. The proposed FDD tool uses a family of multivariate statistical methods called MultiDimensional Scaling to deal with this huge amount of data such as the early work proposed in (Torgerson, 1952). This method is for example used in the field of electrochemical energy storage to detect faults during batteries operation and estimate their life expectancy (Degret et al., 2014). This tool, is part of the data-driven methods, may be used to solve the problem of under-qualified maintenance or operation staff. Indeed, thanks to its ability to reduce high dimensional data to a two-dimensional representation space whose enhancing all the identified faults, this tool can be easily used and understood by most of the in situ agents. The goal of this new FDD technique is to create a user-friendly, autonomous staff oriented tool, which provides real-time results. Furthermore, the possibility of using MDS method to represent data according to severity indicators may suggest the opportunity of fault prediction.

The present study proposes to apply the developed tool directly on data from a real existing tertiary building system to appreciate its capacity to detect some faults.

Method

The developed FDD method can be divided in two main steps. The first one is a *pre-process* stage, where the input data are treated to detect faults. The second one is the *data treatment* during which the dimensionality reduction is performed and the outputs are generated. All these stages, presented and detailed in the following section, are performed in the MATLAB environment because it is robust for data treatment.

Pre-process

In the first step of the method the data are treated to create the *rules space*, whose coordinates are the evaluation of the chosen logical rules on measured parameters. These input data are usual data type for building energy management system (BEMS), such as temperature, power, or boolean. For the purposes of the numerical analysis of a given practical problem,

the data are transformed into dimensionless quantities. Indeed, the floating point arithmetics is built such as the rounding errors are minimal if the computer manipulates the numbers of the same magnitude (Kahan and Palmer, 1979). In addition, the density of the floating point numbers is the highest in the interval $]\,0,\,1\,[$ and it decreases when we move further from this interval. Thus, it is always better to handle numerically the quantities of the order of $\mathcal{O}(\,1\,)$ to prevent serious round-off errors. The original data set is defined in the following form:

$$U \;=\; [\,U_{ni}\,]\,, \tag{1}$$

where $U_{ni} \overset{\text{def}}{:=} U_i\,(\,t_n\,)$ is a measured quantity such as the temperature or electrical consumption, obtained on a time grid, $i \in \{\,1\,,\,\ldots\,,\,N_m\,\}$ with N_m the total number of measured quantities and $n \in \{\,1\,,\,\ldots\,,\,N_t\,\}$ with N_t the total number of time acquisitions.

The data are then treated according to the operation rules defined by the user to detect the faults. The operation rules represent some logical statements that must be observed in a regular operation of the building. A rule is defined by:

$$U_{ni} \;-\; U_{nj} \;=\; U_{\text{lim}}\,, \tag{2}$$

where $(\,i\,,\,j\,) \in \{\,1\,,\,\ldots\,,\,N_m\,\}^2$ and $i \neq j$. Here U_{lim} is a chosen limit value so that a fault is detected when the difference $U_i - U_j$ is above this threshold. A total of N_R rules are created and gathered in the following matrix:

$$\rho \;=\; [\,\rho_{nk}\,]\,, \tag{3}$$

with $k \in \{\,1\,,\,\ldots\,,\,N_R\,\}$ and $n \in \{\,1\,,\,\ldots\,,\,N_t\,\}$. Each element of the matrix ρ is defined by:

$$\rho_{nk} \overset{\text{def}}{:=} \frac{U_{ni_k} - U_{nj_k} - U_{\text{lim}_k}}{\beta_k}\,. \tag{4}$$

where $(\,i_k\,,\,j_k\,) \in \{\,1\,,\,\ldots\,,\,N_m\,\}^2$, $i_k \neq j_k$, $k \in \{\,1\,,\,\ldots\,,\,N_R\,\}$ and $n \in \{\,1\,,\,\ldots\,,\,N_t\,\}$. The element ρ_{nk} is a so-called expert rule, where the results of logic rules are treated in a dimensionless form thanks to the factor β_k. This factor is used as a fault intensity scaling parameter by the fault detection process. The result of the expert rules are then processed with an hyperbolic tangent function to create the severity index SI called σ_{nk} and computed according to:

$$\sigma_{nk} \overset{\text{def}}{:=} \tanh(\,\rho_{nk}\,)\,. \tag{5}$$

Thus, the matrix gathering the severity indexes can be defined:

$$\sigma \;=\; [\,\sigma_{nk}\,]\,, \tag{6}$$

where $k \in \{\,1\,,\,\ldots\,,\,N_R\,\}$ and $n \in \{\,1\,,\,\ldots\,,\,N_t\,\}$. It can be remarked that

$$\sigma \in \mathcal{M}at\,(\,]-1\,,\,1\,[\,,\,N_t\,,\,N_R\,)$$

recalling that N_R is the number of rules and N_t is the number of time step measurements. The SI allows to present the results of Eq. (5) in a dimensionless form in $]\,-1\,,\,1\,[$ regardless of the physical nature of the fault. Figure 1 shows a representation of the SI and how the reading must be interpreted.

Another indicator is created to detect the emergence of multiple faults. The total fault severity is denoted by λ and defined as follows:

$$\lambda_n \overset{\text{def}}{:=} \sum_{k=1}^{N_R} \sigma_{nk}^{\,+}\;, \tag{7}$$

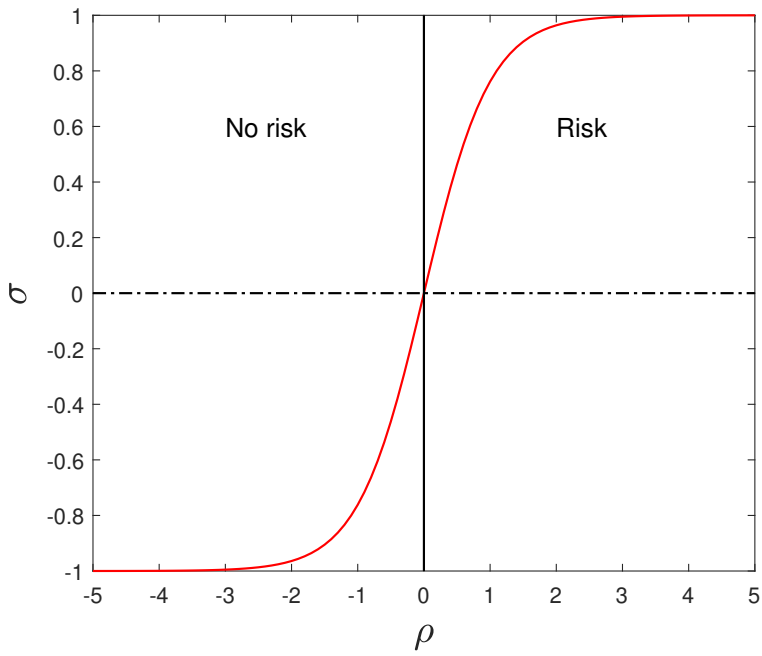

Figure 1: Severity index form.

where $n \in \{1, \ldots, N_t\}$ and

$$\sigma_{nk}^{+} \stackrel{\text{def}}{:=} \max(\sigma_{nk}, 0). \tag{8}$$

Thus, σ_{nk}^{+} corresponds to the positive part of the severity, meaning that a fault appeared as illustrated in Figure 1. The last indicator corresponds to the total number of faults appearing simultaneously:

$$\delta_n \stackrel{\text{def}}{:=} \sum_{k=1}^{N_R} \chi_{[0,1]}(\sigma_{nk}), \tag{9}$$

where $\chi_{[0,1]}(\bullet)$ denotes the indicator function of the subset $[0, 1]$.

Data treatment

The data are treated by the MultiDimensional Scaling method to give information on faults occurrences in the system. The MDS methods uses the data created in the *pre-process* phase as input. The interest of multivariate statistic tools, such as MDS is to represent data of high dimensionality in viewable form, *i.e.* representation in a two or three dimensional space. In this case, the original dimension is a $[N_t \times N_R]$ matrix, where N_t is the number of measurements and N_R the number of rules. For example in our case study, this original space data is constituted by the six rules ρ_k for winter operation. Then, the MultiDimensional Scaling reduces this high dimensional dataset into a lower dimensional representation called a *map*. This map is the most satisfying representation of the original set of data in the new space. To generate this map within a shortest time, the Hastie method (Hastie et al., 2009) is used to identify the most N_H interesting points in the original space. It chooses the point in the original space which minimizes distance to all the other points. Thus, the most representative points of the data set dispersion in the original space are selected.

A MDS method derived from t distributed Stochastic Neighbour Embedding (Maaten and Hinton, 2008) is then used to generate the map, while preserving the point neighbourhood based on the conservation of the distance among the points. So, the MultiDimensional Scaling method is a mapping which for the matrix σ returns a new matrix $\tilde{\sigma}$ in a lower dimensionality space:

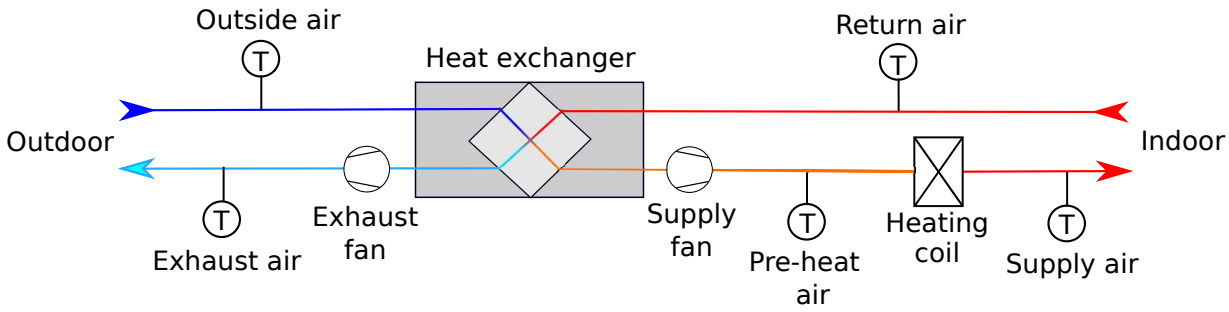

Figure 2: Schematic representation of an AHU.

$$f \,:\, \mathcal{M}at\,(\,]-1\,,\,1\,[\,,\,N_t\,,\,N_R\,) \rightarrow \mathcal{M}at\,(\,\mathbb{R}\,,\,N_H\,,\,N_D\,) \tag{10}$$

$$f \,:\, \sigma \longmapsto \tilde{\sigma}\,, \tag{11}$$

where N_D is the number of dimensions chosen in the reduced space (two or three usually). So, this method allows an easily understandable representation of the initial matrix U of $[\,N_t \times N_m\,]$ size by a matrix of only $[\,N_H \times N_D\,]$. The breaking down of the final map regarding the violation to each rule allows the identification of faulty areas and of nominal operation.

Case study

In this paper, the FDD tool is applied to a French existing building, located in the West of the country. For confidentiality issues, this building cannot be identified in details. The construction is a recently built waste treatment center divided in two parts. The first one is composed of *ca.* 500 m^2 of technical premises and will not be studied in this paper. The second houses the part of the building where the employees work, composed of offices, lunch room, cloakroom and sanitary, on 350 m^2.

The internal conditions of the office part are maintained thanks to a heating floor, radiators and air treatment, all powered by a geothermal heat pump and solar heating network. In this work, the FDD tool is only tested on the air handling unit (AHU). The justification of this restriction will be detailed more later. The AHU is a dual flow ventilation composed of a cross flow heat exchanger (to pre-heat the air thanks to exhaust air), two fans, an heating coil (powered by the heat pump), to regulate the air temperature, and several temperature sensors to see the evolution in the system. Figure 2 illustrates the AHU composition and shows the implementation of sensors.

The system data are temperature informations from the different sensors and power information of fans and the heating coil. The equipment power was turned into boolean values to allow their treatment by the MDS method. The temperatures recorded by different sensors will be called as follows in the rest of the paper : T_{out} , the outside air temperature (read before the heat exchanger) ; T_{prh} , the pre-heated air (after heat exchanger warming) ; as T_{sup} , the supply air (reheated by the heating coil) ; T_{int}, the indoor air (measured in the extraction path) ; and T_{exh} , the exhaust air (at the end of the ventilation path, after the heat exchanger). As for the temperatures, the energy use of the harvest elements will be renamed : the supply fan energy use as P_{fan} and the heating coil consumption as P_{hp} . In this study case $N_m = 7$. The different temperature profiles are presented in Figures 3 to 6.

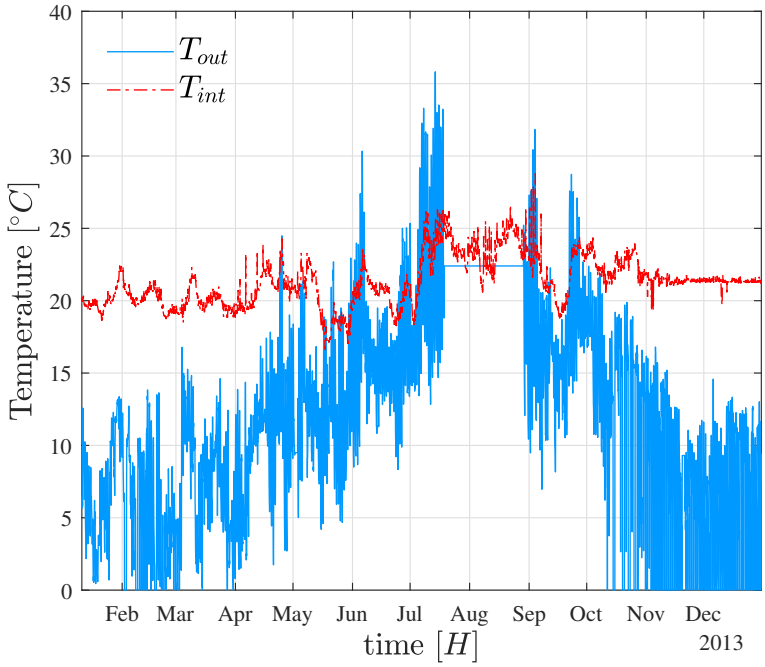

Figure 3: Inside and outside temperature time series.

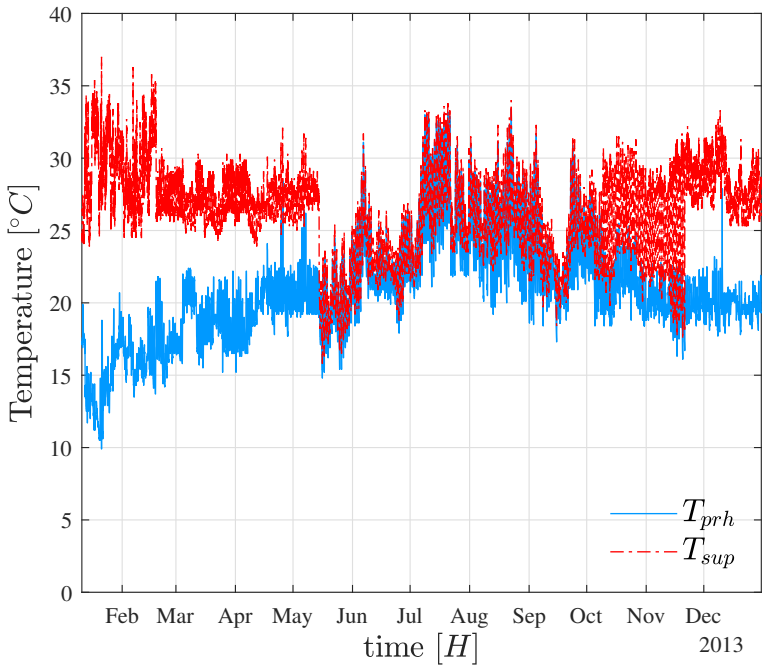

Figure 4: Pre-heat and inlet temperature time series.

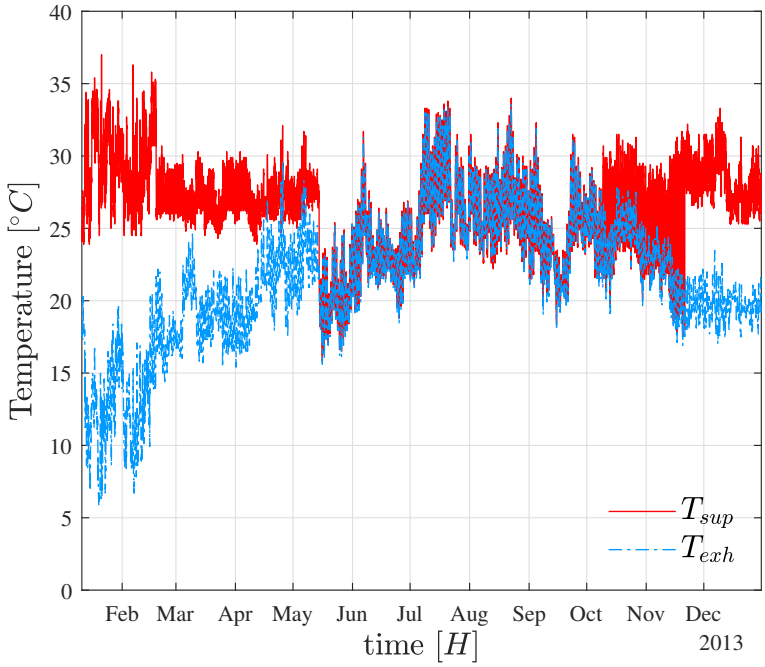

Figure 5: Inlet and outlet temperature time series.

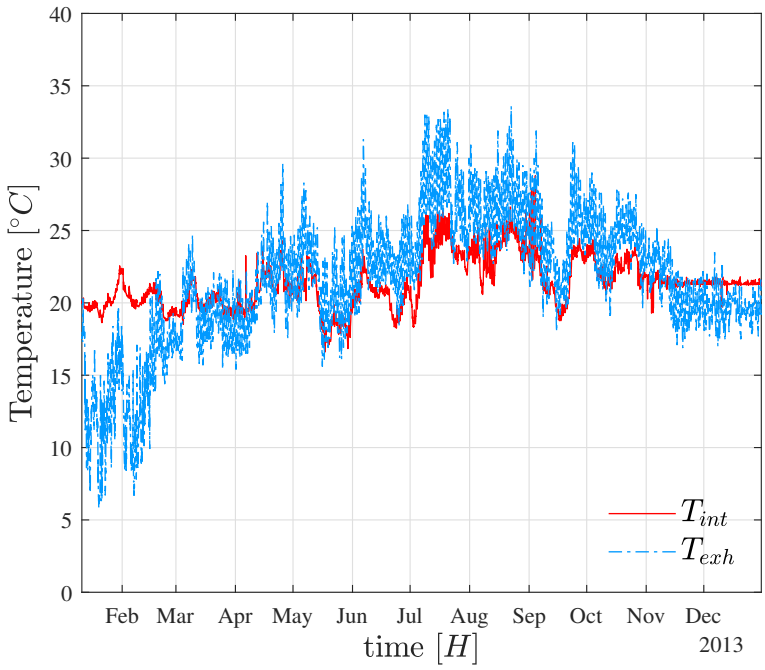

Figure 6: Inside and outlet temperature time series.

Table 1: Set of rules.

Season	Ident.	Rules
Winter	1.1	$T_{\text{sup}} < T_{\text{exh}}$
	1.2	$T_{\text{out}} < T_{\text{int}}$
	2	$T_{\text{int}} + 10 < T_{\text{sup}}$ [°C]
	3	$0 < T_{\text{out}} - T_{\text{sup}}$
	4.1	$T_{\text{int}} < 19$ [°C]
	4.2	$T_{\text{prh}} = T_{\text{sup}}$
	5.1	$21 < T_{\text{int}}$ [°C]
	5.2	$T_{\text{prh}} < T_{\text{sup}}$
	6.1	$T_{\text{sup}} < T_{\text{int}} - 1$ [°C]
	6.2	$P_{\text{hp}} = 0$
Summer	1	$T_{\text{prh}} < T_{\text{sup}}$
	2.1	$T_{\text{out}} < T_{\text{int}}$
	2.2	$T_{\text{int}} < 22$ [°C]
	2.3	$P_{\text{fan}} < 1$
	3	$T_i(t_{n+1}) - T_i(t_n) < 10^{-6}$ [°C]

In this work, since the control sequences of the system are not known, the case study is particularly useful to explore an underrated FDD building area fault : the wrongly configured building system (Lazarova-Molnar et al., 2016). Thus, data values validity regarding several logical control rules will be studied and the ability of this FDD tool to detect these faults will be used as a benchmark.

As the rules are unknown, some must be chosen to determine the *nominal* operation of the system. These control rules have been selected, in order to highlight some operation faults judged plausible from the functional analysis of the building system. The rules chosen to study the operation of the system are derived from the AHU performance assessment rules (APAR) (Schein et al., 2006) and summarized in Table 1.

Six rules are selected for the winter operation ($N_{\text{R}}^{\text{w}} = 6$) and three for the summer ($N_{\text{R}}^{\text{s}} = 3$). Regarding the winter rules, the first rule is composed of two sub-rules, allowing to see if the room is overheating. The second rule shows a control problem with a useless overheating of the air. The third one is used to detect a failure in heating components or bad sensors implantation. The set of rules n°4 is used to know if the coil did not operate when it is needed. The rules 5.1 and 5.2 detect if the coil operates when not necessary. And finally the two rules n°6 show if the heating pump provides energy only when it is needed. Concerning the summer season, only three rules are proposed. The first one checks if the heating pump operates in a bad regime (*i.e.* heating mode), the second one if the free-cooling is activated and the third checks if the temperature sensors are saturated.

Then the U_{lim} and β values must be set regarding the goal, to allow the data treatment. These two coefficients must be selected by the operator for each rule presented above. They are issued from expert knowledge and their determination may be complex. The values of these two coefficients are presented in Tables 2 and 3 for the winter and summer periods, respectively.

All this information must be detailed for the MDS treatment, since it is the basis for the dimensionality reduction. After the selection and the code implementation of all these

Table 2: U_{lim} and β values for winter.

Ident.	1.1	1.2	2		3	
U_{lim}	0	0	-5		0	
β	2	5	10		5	
Ident.	4.1	4.2	5.1	5.2	6.1	6.2
U_{lim}	19	0	-21	0	-1	0
β	1	1	2	1	1	0

Table 3: U_{lim} and β values for summer.

Ident.	1	2.1	2.2	2.3	3
U_{lim}	-2	0	-22	0.75	10^{-6}
β	1	1	1	1	$\frac{\left(T_i(t_{n+1})+T_i(t_n)\right)}{2}$

parameters, the dimension reduction with MDS is carried out.

Results and Discussion

In this section we present the results obtained using this method, first the values generated in the *pre-process* stage and then the map is obtained in the *data treatment* phase.
Regarding the *pre-process* stage, the input matrix U has a size of $[8512 \times 6]$, the ρ and σ are matrices of size $[4346 \times 6]$ and the $\tilde{\sigma}$ matrix has a size of $[1000 \times 2]$ for the winter season. For the results of the *data treatment* stage, the map generated by the reduction method reflects the behaviour of the system regarding the operational rules described above. The preliminary results are shown in Figure 7 for the winter period. On this representation of the system operation faults, each point represents the state of the system at each time step. On one hand, the color of the marker indicates the total fault intensity. The value shown on the right with the color bar is the results of λ *i.e.* the sum of positive parts of the severity indicators σ_k above zero. This value ranges from 0 with a black marker, meaning that the resulting value of the expert rule are under the prescribed limit, up to 2 (white markers) indicating simultaneous faults (two rules severity indicators reach the maximal value of one, or more faults reaching intermediate positive values). A value of the SI of approximately 1 (orange markers) means that there is a complete violation of one of the rules or a partial violation of at least two rules. On the other hand, the marker shape shows how many faults are effectively appearing simultaneously. A round marker indicates that there is no severity indicator above zero, a square marker means that one SI is above zero, two are represented with a diamond and three with a triangle though this case is extremely rare in the data set. So, on this map it cannot appear a round marker with an other color than black, or a square marker with a lighter color than orange, because the number of potential faults would not match with the λ value.

So, the analysis of the projected data shows that the nominal operation of the system is essentially concentrated in the middle of the map. Indeed, all square and blue markers are clustered in the central area and it is obvious that a continuum leads to different faults located on boundary of the map. It can also be noticed that even though square and diamond markers

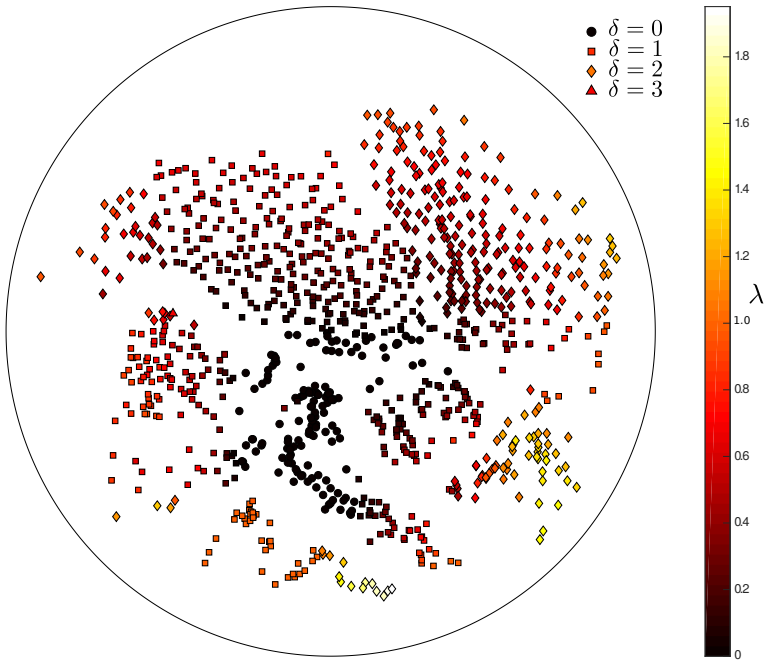

Figure 7: Air Handling Unit winter operation map generated by MDS.

may be close, they are significantly clustered. These fortuitous attributes of the map generated by the MDS method, provide a clear representation and identification of faults. Actually, since the nominal operation is located in the center of the map, with the identified faults clustering on the boundary, their determination is easy and quick.

The next step is the identification of the different kinds of faults. The expected result of the method, is that all the faults are clearly clustered, and to avoid a dispersion of the same faults across all the map. An identification of the violated expert rules leads to this new representation of the map of the faults repartition shown in Figure 8. In this map each color as well as the shape of the markers, represents one of the six winter rules. A point is identified as displaying the fault ρ_i if it has the highest severity indicator value σ_i. So the representation of the point (color or form) indicates the predominant fault. Figure 8 also represents a possible view that may be displayed to an operator on the BEMS.

Thanks to the color code, the visualisation of the faults' location in the reduced space is rapidly understandable. The analysis of this map regarding the faults identification highlights the fact that the method clusters the faults together. This capacity of the dimensionality reduction method is particularly appreciable for the aims of an Automated Fault Detection and Diagnosis (AFDD) tool. Indeed, the prior data treatment coupled with the MDS reduction method allows the generation of a new reduced space, where all the identified faults are clearly grouped. The analysis of the Figures 7 and 8, shows that the evolution of the operation of the system appears to be continuous. The condition of nominal operation, located in the center of the map drifts to faulty behaviour, which are rather located on the map boundary. The direction of the operation of the system representation indicates which fault type is overriding. Simpler, on the reduced space, the radius length of the localisation can be understood as faults gravity and the angle type of the occurring fault. This representation is user oriented, because it allows a quick and easy identification of the faults.

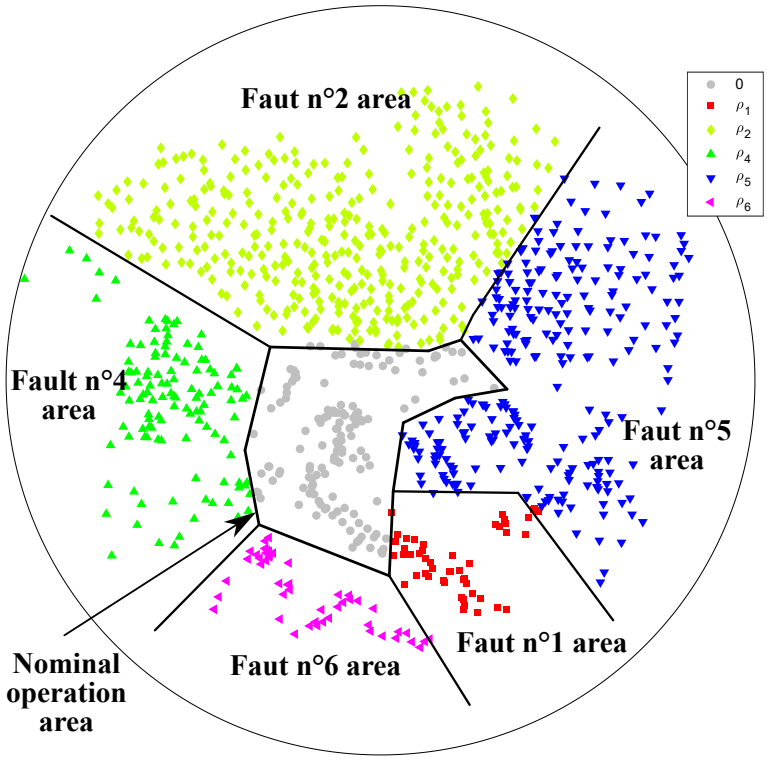

Figure 8: Main fault identification for the winter period.

Thanks to the two different representations of the markers in the above maps, it is straightforward to identify advantages of the this kind of representation for FDD application. Another asset of this FDD tool, is that as Figure 7 already showed, it allows multiple faulty states detection (markers in diamond or triangle shapes). With Figure 9 map representation, this multiple fault detection can be more extensively studied. In this new map, as in Figure 8, each different marker shape stands for the principal winter fault regarding the severity indicator result. In addition, the marker color indicates the second most significant fault, *i.e.* the fault with the second most important SI value. The grey points represent the points with at most one fault.

This analysis of the map in Figure 9 reveals that the point of each time step in the new space is also logical regarding the multiple fault occurrences. Indeed, as clearly illustrated in the top right corner area of this map, once again the MDS allows the creation of a space continuum between the fault evolution, in which the multiple fault appears. In this aforementioned zone, there are two main faults which are the fault ρ_2, symbolised by the diamond marker, and the fault ρ_5, represented by the apex down triangle. In Figure 8 we can only see that the boundary of these two faults is located in a common region, with some almost overlying points. However, thanks to Figure 9, these layering points are explained. Actually, this map shows that in the triangle nearer part of the diamond zone a multiple fault appears, with the fault number two as second fault (showing by the apex down triangle as marker) illustrated by the red color. And in the same way, in the apex down triangle zone, closer area of the diamond zone, appears triangle in cyan color. Near this *overlapped zone* there is an area of multiple faults and these overlying points can be explained by the fact that in this zone the main fault changes from one to another.

The operation map for the summer period is shown in Figure 10. As mentioned before, the nominal operation is located in the map center. Figure 11 enables to identify the main fault occurring. On the bottom boundaries, the three main faults occurrences are highlighted. The

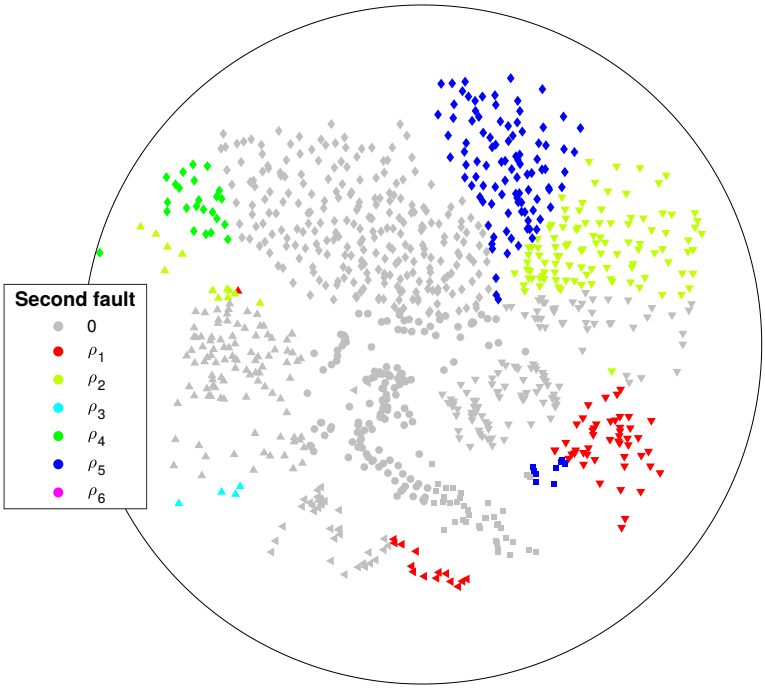

Figure 9: Combined fault identification for the winter period: second fault appearing after the main one.

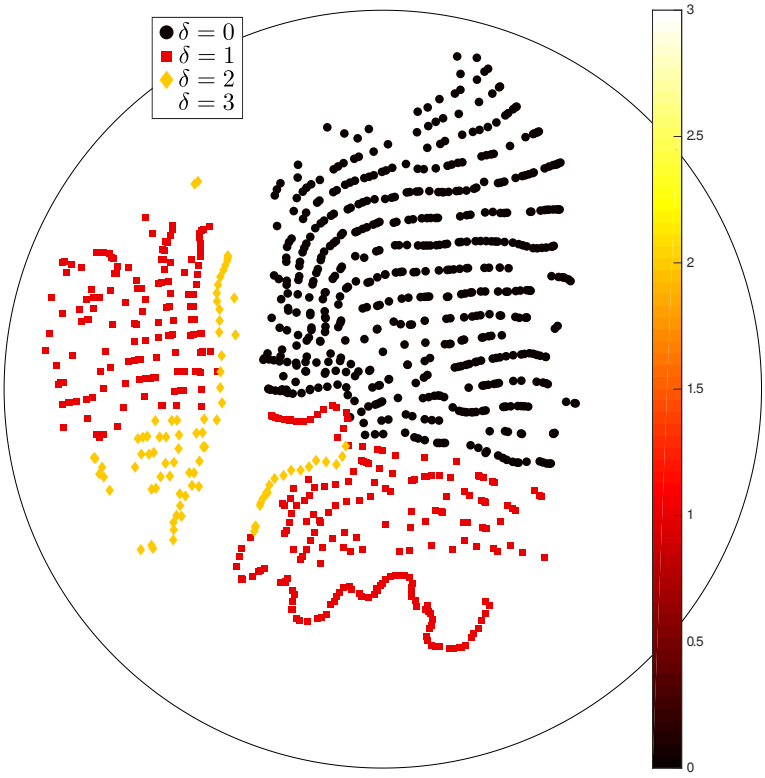

Figure 10: Air Handling Unit summer operation map.

Figure 11: Main fault identification for the summer period.

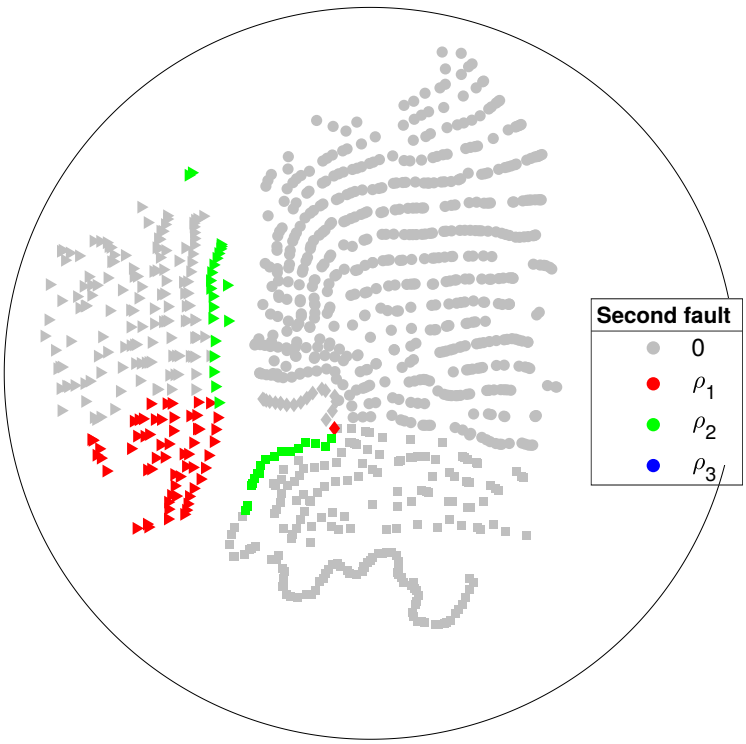

Figure 12: Combined fault identification for the summer period: second fault appearing after the main one.

faults n°1 and 3 are more frequent than faults n°2 . It can be noted that the saturation of the sensor of the outside temperature is well identified using fault n°3 . The saturation can also be noticed in Figure 3 during the month of August. When two faults occur simultaneously, they can be distinguished in Figure 12. By comparison between Figures 11 and 12, in the bottom left corner, the main fault corresponds to the rule n°3 and the secondary to rule n°1 .

Regarding the Central Processing Unit (CPU) time of this method, it is relatively reasonable comparatively to the engineering time needed to analyse all these data and to extract information. This cost is almost entirely due to the MDS mapping construction. The elaboration of a map of about 1000 points as presented in this study, represent around 20 min , measured on a computer with Intel(R) Core(TM) i7 − 7820HQ and 32, 0 Go RAM and the algorithm complexity is N_H^2 with N_H the number of plotted points.

We highlight some further interesting and promising capacities of the developed intuitive FDD tools. A good definition of the severity index seems to be potentially used to make some fault prediction. This flexibility may be contained in the selection of the two crucial values of U_{lim} and β. These parameters are not trivial to set, and must be chosen carefully by an expert. The fault prediction should be studied more deeply regarding these two parameters.

Conclusion

This study shows that the MultiDimensional Scaling provides an interesting and promising methodology for air handling unit Fault Detection and Diagnosis, but more generally for systems. Another interesting area of the research is the possibility to play with the two problems in dimension reduction between *tears* and *false neighbourhoods* in the map. Indeed, these two inevitable consequences of dimension reduction methods can be used to increase the efficiency of our FDD approach and more specifically the *tears*. This perspective of using this inherent simplification of the reduction method must be an asset for the usage of the multivariate statistics for Fault Detection and Diagnosis.
However the study should be continued further to check the robustness of the tools with, for example, data sets which contain even more than three simultaneous faults. Also, as the tool is tested on real data from an anonymous building, the authors miss feed back on the actual faults, so the tool must be evaluated on more controlled experimental data and thus estimate precisely the probabilities of false positives or negatives. Further investigations should be also made to allow the creation of real automated fault detection tools, as for the identification of the faulty area definition, the faults apparition, *etc.* And the CPU time should be also reduced even more to have a real-time FDD tool, so more work is needed to reach this aim.

Acknowledgements

We thank the Cerema team for their participation in this study. This project is supported by the Interreg V France-Switzerland European Territorial Cooperation Program and has benefited from a European grant of 607776.49 € through the European Regional Development Fund (ERDF) as well as federal funds Interreg Switzerland for150001 CHF and 123999 CHF for cantonal and communal aid. The authors acknowledge the Junior Chair Research program "Building performance assessment, evaluation and enhancement" from the University of Savoie Mont Blanc in collaboration with The French Atomic and Alternative Energy Center (CEA) and Scientific and Technical Center for Buildings (CSTB).

References

Bruton, K., P. Raftery, B. Kennedy, M. M. Keane, and D. T. J. O'Sullivan (2014). Review of automated fault detection and diagnostic tools in air handling units. *Energy Efficiency 7*(2), 335–351.

de Wilde, P. (2014). The gap between predicted and measured energy performance of buildings: A framework for investigation. *Automation in Construction 41*, 40–49.

Degret, F., A. Petitjean, C. Primot, M. Alias, N. Guillet, S. Lespinats, P. Maillet, and P. Thivel (2014). Emission Acoustique : Indicateurs d'état Des Batteries Lithium Ion. In *Groupe Français d'étude Des Composés d'insertion, GFECI 2014*, Ile d'Oléron, France.

Djuric, N. and V. Novakovic (2009, February). Review of possibilities and necessities for building lifetime commissioning. *Renewable and Sustainable Energy Reviews 13*(2), 486–492.

Hastie, T., R. Tibshirani, and J. H. Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer series in statistics. New York, NY: Springer.

(1999). *Annex 25 : Real Time Simulation of HVAC Systems for Building Optimisation, Fault Detection and Diagnostics*.

Kahan, W. and J. Palmer (1979). On a proposed floating-point standard. *ACM SIGNUM Newsletter 14*, 1321.

Katipamula, S. and M. Brambley (2005). Review Article: Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems—A Review, Part I. *HVAC&R Research 11*(1), 3–25.

Kim, W. and S. Katipamula (2018). A review of fault detection and diagnostics methods for building systems. *Science and Technology for the Built Environment 24*(1), 3–21.

Lazarova-Molnar, S., H. R. Shaker, N. Mohamed, and B. N. Jorgensen (2016). Fault Detection and Diagnosis for Smart Buildings: State of the Art, Trends and Challenges. In *2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC)*, Muscat, pp. 1–7. IEEE.

Li, Y. and Z. O'Neill (2018). A critical review of fault modeling of HVAC systems in buildings. *Building Simulation 11*(5), 953–975.

Maaten, L. v. d. and G. Hinton (2008). Visualizing data using t-SNE. *Journal of machine learning research 9*(Nov), 2579–2605.

Schein, J., S. T. Bushby, N. S. Castro, and J. M. House (2006). A rule-based fault detection method for air handling units. *Energy and Buildings 38*(12), 1485–1492.

Torgerson, W. S. (1952). Multidimensional Scaling: I. Theory and Method. *Psychometrika 17*(4), 401–419.

Yu, Y., D. Woradechjumroen, and D. Yu (2014). A Review of Fault Detection and Diagnosis Methodologies on Air-Handling Units. *Energy and Buildings 82*, 550–562.