

# ISC2 Technical Report

Towards a Methodology for the Characterization of IoT Data Sets of the Smart Building Sector

April, 2022

## Abstract

Here we describe the procedures and experimental results of the paper Towards a "Methodology for the Characterization of IoT Data Sets of the Smart Building Sector." We also list an extended version of the "Related works" section and bibliographic references helpful in context. In short, we provide more technical material, motivations, and context elements here than in the paper. The paper aims to concentrate on the approach's results and findings.

## 1. Introduction and problem description

Authors in (11) introduce the vision of autonomic computing—computing systems that can manage themselves given high-level objectives from administrators. As noted in (12), the term autonomic computing is *symbolic of a vast and somewhat tangled hierarchy of natural self-governing systems. Many of which consist of myriad interacting, self-governing components that comprise large numbers of interacting, autonomous, self-governing components at the next level down.* We propose, in this paper, to consider a smart building (its digital manager) as an autonomic system because the concept corresponds to what the building manager wants. For instance, he does not plan to allocate people for manual or semi-automatic tasks in the building to regulate the temperature of the rooms.

From an architectural point of view, see Figure 1, an autonomic element will typically consist of one or more managed elements coupled with a single autonomic manager that controls and represents them. By monitoring the managed element and its external environment and constructing and executing plans based on an analysis of this information, the autonomic manager will relieve humans of the responsibility of directly managing the managed element. This view is the so-called IBM MAPE-K model. It corresponds to the idea that we build knowledge over the managed element through monitoring, analyzing, planning, and executing sequential steps.

The work of R ve de Sc nes Urbaines (RSU) (7), Plaine Commune's industrial demonstrator located in the north of Paris, France, focuses on building a sustainable, inclusive, but also adaptable city in an existing urban fabric. Regarding the ecosystem of smart buildings, among the 47 ideas developed by the demonstrator and already pre-selected by the RSU community, we can cite the

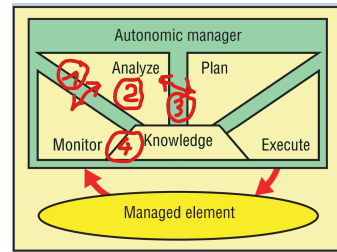


Figure 1: The MAPE-K model

project of allowing the reversibility of buildings - first Olympic housing, then family accommodation whose size can vary over the years and, perhaps in the future, offices-, imagining new urban forms - such as barges on the Seine river-, fighting against heat islands, developing a concrete capable of accommodating plant species, building a 3D concrete manufacturing plant, etc.

The scenario we investigate in this article is as follows. A particular actor, private or public, has the technical responsibility of managing a smart building. Moreover, he has to respect a corporate social responsibility policy, especially concerning the issues of digital sobriety. This actor uses a BIM-like tool for modeling. If the building is already operating, it is eventually controlled by a smart building manager.

A BIM designates the tools for modeling the construction information implemented by applications that allow modeling the data produced in the building, a structure, or an engineering structure, i.e., construction of great importance and size. It, therefore, has information about the locations and types of sensors deployed in the building, for example, temperature, humidity, and CO2 sensors. The data produced by the sensors can be sent to a cloud or kept at the edge of the network.

In this context and to go towards more digital

sobriety, we provide the building manager with a methodology to launch analysis on the sensor data and then build a smart building model. Finally, based on the previous automatic studies, we sketch some plans to reduce data emissions from the sensors while preserving the model. This part of the work is not deduced automatically in this paper for simplicity. The reader should notice that we are working on the analysis plan of the MAPE-K model (see symbol ② on Figure 1) and also on the interfaces such as cleaning and planning (see symbol ① on Figure 1) and predicting (see symbol ③ on Figure 1). We aim at generating knowledge for the element (see symbol ④ on Figure 1). But again, the core of the paper is on the analysis plan.

Since a fully automatically managed building, in the sense of the MAPE-K model, is currently out of reach, the aim is to provide a decision aid, enabling the skilled expert to make informed choices about data production.

Monitoring buildings is crucial for many aspects, including mastering heating costs. The U.K.'s housing stock is particularly leaky. After five hours, the temperatures in British houses drop about 3° Celsius, the highest rate of countries in western and northern Europe surveyed by thermostat company tado GmbH. In Germany, the drop was only 1° over the same time. Check the footnote<sup>1</sup> for a panorama of heating leaks in Europe. Moreover, we know that over 17% of all carbon pollution in the U.K. comes from buildings. The (green) heating resource would be better utilized to charge an electric bike or a shared scooter than to escape from a window.

The Internet of Things (IoT) domain has been one of the fastest-growing areas in the computer industry for the last few years (13). The global sensor market is large and growing fast. By one estimate, it is projected to reach U.S. \$346 billion in sales by 2028, up from \$167 billion in 2019 (6). Consequently, IoT applications are becoming the dominant workload for many end-to-end systems such as clouds, fog, or edge computing systems (9). Knowing the nature of expected workloads may help build efficient systems to store and analyze data and guide exploration of the data. Thus, it is fundamental to characterize IoT workloads, but to the best of our knowledge, only a few studies are related to that topic. Moreover, if we aim to also learn about the data sets, we would obtain significant insights for mastering a sober Internet. For instance, a good motivation in this context of sober computing is the reduction of communica-

tions from sensors if the sensors have learned that data is or is not helpful for the intended application. Throughout an empirical study, we provide recommendations of this sort and are motivated by an upstream analysis of our analyzed data sets.

Among the many supplementary questions in analyzing, characterizing, and understanding the data sets, we would like to mention the following addressed in the paper. Are there any anomalies in the data sets? Are they to be taken into account for the analyses? How do we combine anomalies detected by several algorithms? Are the two studied data sets correlated in some way? For instance, can Spatio-temporal data identify aggregated data, i.e., heterogeneous data? Can the data quickly be clustered, and how many clusters can be identified?

### 1.1. General approach

The system we consider in this work has IoT devices that send data at irregular intervals, independent of each other. The nature of data can be binary data, strings, timestamps, reals, or integers, and some data sets present metadata.

The methodology we developed in this paper, and the organization of the paper, follow the MAPE-K model for the production of knowledge (see symbol ④ in Figure 1). Thus, we first aggregate, clean, and filter the data (see symbol ① in Figure 1). Then we put in place analyses (see symbol ② in Figure 1). At last, we generate semi-automatic plans (see symbol ③ in Figure 1). All of this constitutes the methodology we promote in this paper.

Our methodology for analyzing data consisted of the following steps:

1. Eventually, the data are aggregated (see the discussion below) from all the files that make up the data set;
2. The possibly aggregated data set is cleaned by removing any invalid and duplicate data;
3. The data are filtered based on information such as days, hours, month and so on;
4. The filtered data is characterized by computing event counts and statistics. Data is combined through machine learning algorithms such as anomaly detection and classification (clustering). This step aims to provide multiple views of the data to infer expertise about the building or knowledge about the collecting system. *This double aspect concerning both the building and the information gathering device is the particularity of our work to build a piece of business knowledge and skills;*

<sup>1</sup><https://www.tado.com/gb-en/press/uk-homes-losing-heat-up-to-three-times-faster-than-european-neighbours>

Let us say a word on the fusion of data. We could merge our sources, which can lead to inconsistency problems. For example, is it possible to combine Co2 data taken outdoors and Co2 data taken indoors? To avoid explaining an axiomatization of both the modeling rules of the inconsistencies, how to detect inconsistencies, and finally, how to clean up inconsistencies, we propose fusion data in a "natural" way. In the end, the number of attributes of the data tables is not so high.

We will therefore be able to analyze the data sets separately, then consider merging these same data sets into a new data set. With this new augmented data set, the general idea is to check whether we are or not in the presence of new facts.

In the case of time series, such as energy consumption data, completeness, outliers, timeliness, and accuracy are among the most critical dimensions of the problem. Completeness measures whether some data are missing, accuracy measures whether the samples are correct and reliable, and timeliness measures whether the information is up to date. At the same time, outliers are inconsistent data with the rest of the series, e.g., a statistical outlier.

After cleaning, this paper mainly focuses on the outliers detection and classification problems. Indeed, we observed some messages in the raw data coming from the server used to query the data, hence the cleaning action. Moreover, the completeness analysis is complex because some data may be shifted in time and not missing due to the drift of the sensors' clock.

## 2. Data sets, Method and Equipment

### 2.1. IoT data sets

All the two data sets we deal with in the paper come from a building located in France. These two data sets have been built, by others, by concatenation of data with the granularity of 1 minute, using instant values for DS-2 and aggregated data for DS-1. We do not master this part of the production. For instance, we have no idea about the principle used for the aggregation. At the moment, the two data sets are not publicly available for privacy concerns, but anyone can use a public API<sup>2</sup> to extract data and build his own data set. However, our codes for analysis are made publicly<sup>3</sup>, and everyone can run them on his data set.

The building has more than 22,000 square meters of floor space divided over six floors and the roof. It is a massively monitored and controlled

building with more than 1,500 sensors, including about 330 electrical energy meters. The measured data is used to monitor internal conditions and to track consumption. We use, for our experiments, a first sub-data set that corresponds to the monitoring of the building from Thursday, Dec 16, 2021, to Friday, Jan 7, 2022. This first sub-data set (**greener-dashboard-20220107.csv**) has 74 attributes (without the timestamp), ranging from the current electricity consumption to the wind direction.

For our analysis, we keep the following attributes because they almost all correspond to metrics of "fluid." Indeed, we assume that our study makes a focus on the fluids:

- **CO2\_max**: maximal instantaneous value observed, overall rooms, for the CO2;
- **CO2\_min**: minimal instantaneous value observed, overall rooms, for the CO2;
- **CO2\_mean**: mean instantaneous value observed, overall rooms, for the CO2;
- **electricity\_daily\_current**: current daily electricity consumption;
- **electricity\_weekly\_current**: current weekly electricity consumption;
- **weather\_tempCurrent**: current outside temperature;
- **weather\_tempMin**: minimal temperature over one day;
- **water\_daily\_current**: the amount of water consumed in the last 24 hours;
- **water\_weekly\_current**: the amount of water consumed in the last week;
- **wind\_peak**: maximum wind speed (wind chill);
- **wind\_speed**: current wind speed;

The second sub-data set, stored in CSV format into the **greener-comfort-20220107.csv** file, contains CO2 and temperatures measures as well as the room's name where sensors are located and the timestamps for each measurement. The data set has 351 sensors for the temperature and 76 sensors for the CO2. All CO2 sensors are located in rooms with a temperature sensor. We found two CO2 sensors with outliers. The sensors have been canceled for the graphics. The date range for the data is also from Dec 16, 2021, to Jan 7, 2022. Each data set has approximately 38k lines. We can assert the data are spatio-temporal.

<sup>2</sup><http://mhi-srv.g2elab.grenoble-inp.fr/API/>

<sup>3</sup><https://gricad-gitlab.univ-grenoble-alpes.fr/batpred/compsac2022>

For our cross-analysis, we keep the following attributes because they all correspond to metrics of "weather." Indeed, we assume the weather has a lot of impact on inside comfort by these features:

- **sun\_direct**: direct light received by the weather station above the building;
- **sun\_diffuse**: diffuse light received by the building ;
- **sun\_global**: total luminosity received by the building ;
- **weather\_tempCurrent**: current outside temperature;
- **weather\_tempMin**: minimal temperature reached during the last 24 hours;
- **weather\_tempMax**: maximal temperature reached during the last 24 hours;
- **wind\_peak**: maximum wind speed (wind chill);
- **wind\_speed**: current wind speed;
- **wind\_direction**: current wind direction, following a 360 degrees circle starting north and increasing clockwise ;

## 2.2. Naming conventions

The data sets naming conventions are:

1. Data set 1, namely DS-1, stands for the **greener-dashboard-20220107.csv** file;
2. Data set 2, namely DS-2, stands for the **greener-comfort-20220107.csv** file;

## 2.3. Outline

The methodology we follow in this section for the analysis step is as follows. First, notice that we conducted different analyses to illustrate the directions of our work, and for the sake of conciseness, we did anomaly detection and classification. Second, we successfully applied a series of treatments on the CO2 attributes present in the two data sets to infer knowledge regarding the room with the most and the minor problem with CO2 issues. Third, we examine the hierarchy empirically, given a certain amount of clusters.

The main goal of the anomaly detection analysis is to identify the observations that do not adhere to general patterns considered normal behavior. In the literature, several Machine Learning (ML) algorithms have been proposed to detect anomalies, each one returning anomalies, potentially different, according to the particular mechanism used in the search process. Hence, it is difficult to say that a given algorithm is better,

in general than another one in finding anomalies for particular data sets. For the anomaly detection algorithms, the portfolio of algorithms that we have used in the experiments is composed of three "orthogonal" algorithms regarding the techniques for the detection, and widely used in the literature, namely iForest (Isolation forest) (15), LOF (Local Outlier Factor) (5), and the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) (8) algorithms.

Cluster analysis is a task that concerns itself with the creation of groups of objects, where each group is called a cluster. Ideally, all members of the same cluster are similar to each other. Equivalently, they are as dissimilar as possible from objects in another cluster. There is no single definition of a cluster, and the characteristics of the objects to be clustered vary. Thus, there are several algorithms to perform clustering. Each one defines specific ways of expressing what a cluster is, how to measure similarities, find groups efficiently, etc. To provide an overview of the behaviors clustered, we perform hierarchical clustering over the time series of DS-1. According to an agglomerative policy, clusters will grow and fuse until one cluster contains all sensors. To processes time series, the algorithm will compute a point-wise metric over the whole length of the dataset, timestamp per timestamp. It is equivalent to comparing the shape of curves to cluster them. We tried two metrics. 'single' allows comparison based on the Euclidean distance between values for each timestamp. 'ward' uses the ward variance minimization algorithm to focus on cluster density. The Ward method strongly refers to the idea of similarity inside clusters and creates more separable clusters, so we chose it to expose different behaviors in-room sensors.

We adopted the Support vector machines (SVMs) algorithm (4) for the classification algorithm, a set of supervised learning methods used for classification, regression, and outliers detection. The advantages of support vector machines are that it is effective in high dimensional spaces and still effective in cases where the number of dimensions is greater than the number of samples. Moreover, SVM uses a subset of training points in the decision function (called support vectors), which is also memory efficient. At last, different Kernel functions can be specified for the decision function. Standard kernels are provided with well-established tools (Scikit-learn, R...), but it is also possible to define custom kernels.

To summarize, since we have no apriori knowledge about the data, the methodology we propose to observe the IoT data concerns multiple anomaly detection algorithms, one time series clustering al-

gorithms, and multiple classification models. We cover various views that characterize our data, and each view corresponds to a particular point of view on the data. We hope that this thorough approach offers orthogonal views to analyze the data. The methodology is divided into two parts. The first part aims to get quick feedback on our data regarding anomalies and clusters. Eventually, we can 'clean' or focus on specific data for the second step. The second part is dedicated to learning about the raw or specialized data issue from the first step through classification.

#### 2.4. *Equipement and Software*

Technically speaking, we implemented and tested some parts of the analyses with R (RStudio-2021.09.2) on Ubuntu 21.10 for Linux PC or Raspberry pi, or MacOS 12.1 Monterey. In this case, the package dependencies are dbscan, DDoutlier, solitude, cluster, clv for the machine learning algorithms, Rtsne, tidyverse, and rio, RColorBrewer, and knitr for the visualization. The complementary work is with Python 3.8.8 with Panda and Scipy.

### 3. Results

We recall and develop here the links enlighten by our study between data and the environment of the building, inside and outside.

#### 3.1. *Results of DS-2 study*

General conclusions from the initial observations and analyses over DS-2 include the feasibility of behaviors clustering and inducted expectations. We separately applied clustering algorithms to CO2 and temperature features to notice rooms with similar behaviors. Figure 2 shows the dendrogram and the numbers of clusters along the hierarchy for the CO2 feature alone. The exact process is applied to cluster temperature behaviors along with the rooms of the building, granting the Figure 3. CO2 measures lead to cluster nearby rooms more often than temperature measures, in figures 2(a) and 3(a). Temperature distribution over the building may vary more than CO2, which explains more temperature sensors. The pairing we notice still questions the use of 74 CO2 sensors and shows that information may be deducted from one sensor instead of two, allowing fewer sensors.

As mentioned before, we provide intelligence for data with little expert knowledge of the building. Using weather features allows us to produce new dendrograms to offer insight into the dependencies of indoor comfort on the weather, providing Figures 4 and 5. We emphasize that the weather has little or no impact on indoor CO2 for

our studied building. Weather metrics evolution is not related to indoor CO2 change and behaves very differently, as shown in Figure 6. The clustering of CO2 sensors displays tree well separable behaviors while the temperature sensors clustering displays Four clusters, one of which is subdivided into 3. On-site metadata and further observations of the building could explain them. As figures 4(b) and 5(b) suggest, small clusters aggregate fast. They are all similar, so it would be hard to find why they are grouped. Figure 4(b) shows irregular plateaus that may hint at separable behaviors on smaller clusters when they are still 5, 13, 17, or even near 30 clusters left, that new metadata like occupancy or architecture elements may explain. The smoothness of the curve on Figure 5(b) only suggests that explanations would be exponentially hard to define, as we need to combine more accurate expert knowledge and metadata.

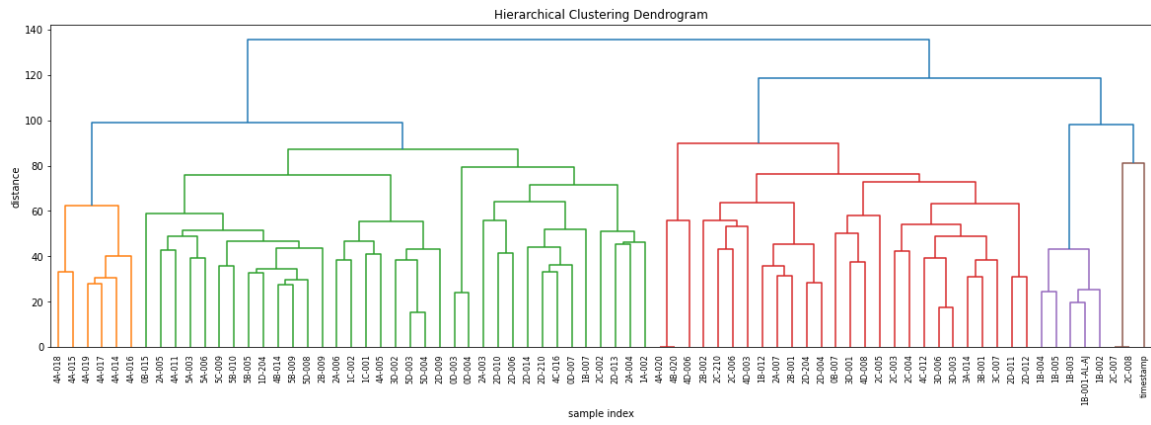
Indoor temperature clustering highlights building architecture and weather data connections to indoor comfort. We notice an overrepresentation of 4th and 5th grounds rooms behaviors in the clusters shown in Figure 7 and a clear separation of ground floor behaviors in Figure 8. Figure 9 highlights the impact of sunlight on some rooms temperatures. From our knowledge, this part of the building faces the South/South-Est directions, which is relevant to high exposure to the Sun.

### 4. Related work

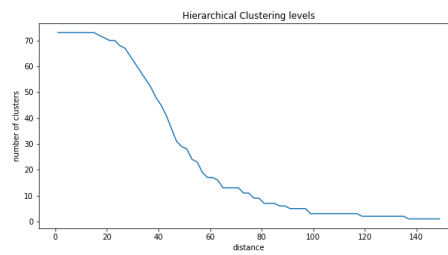
This section relates relevant studies for the characterization of IoT data sets. We assume the reader is familiar with all the referenced classical machine learning algorithms used in this paper.

In (1) authors deal with a generic but straightforward way to model the workload of typical IoT applications to obtain a practical and reproducible method to emulate loads for data centers. The authors combined a reference architecture and one IoT benchmark. The IoT benchmark populated the functional areas of the IoT reference architecture with benchmark components to emulate their typical computing and communication requirements behavior. In short, the paper's originality is in using a category of IoT benchmark, which is neither empirical nor mathematical. Contrary to this work, ours does not provide a realistic load generator, but we analyze real IoT traces. Our data could serve as a new user story under the paper's framework.

In (21) authors described a methodology to analyze IoT traces, described and analyzed three publicly available IoT data sets, and presented a capacity planning study based on the characterization of IoT traces. Contrary to this work, ours does

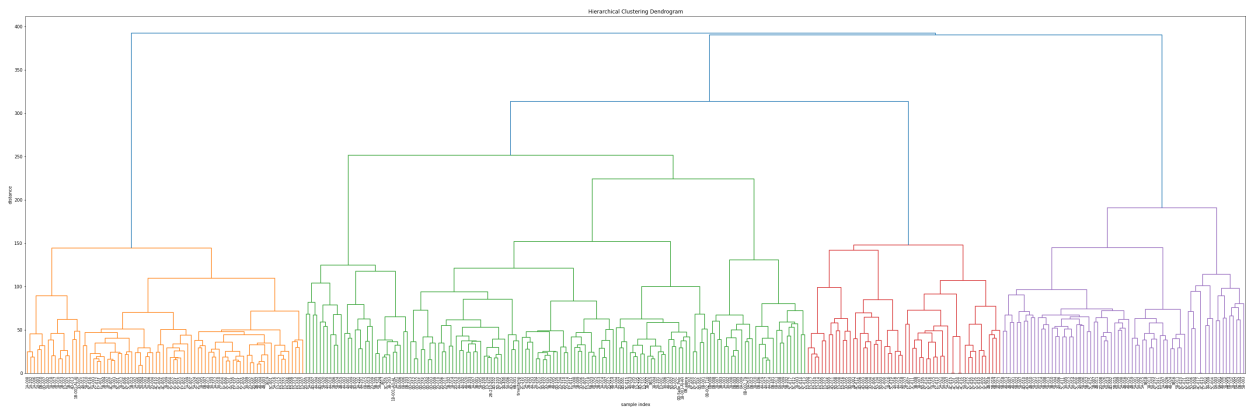


(a) CO2 clustering dendrogram

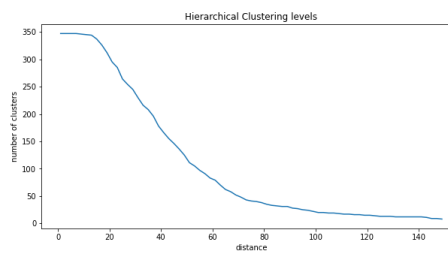


(b) Number of clusters along hierarchy

**Figure 2: CO2 sensors clustering**

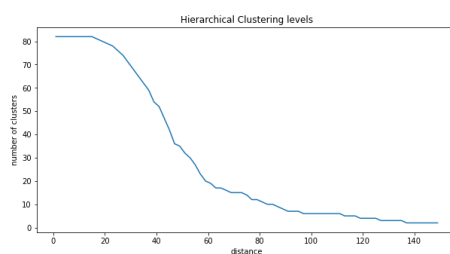
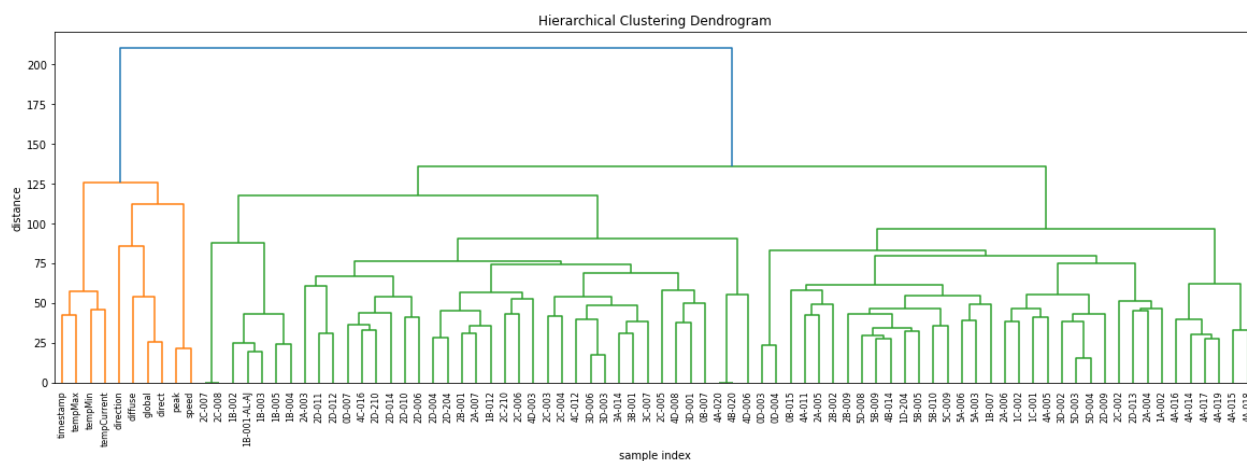


(a) Temperature clustering dendrogram

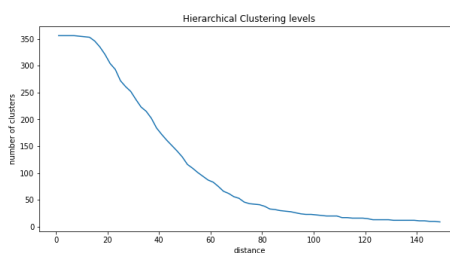
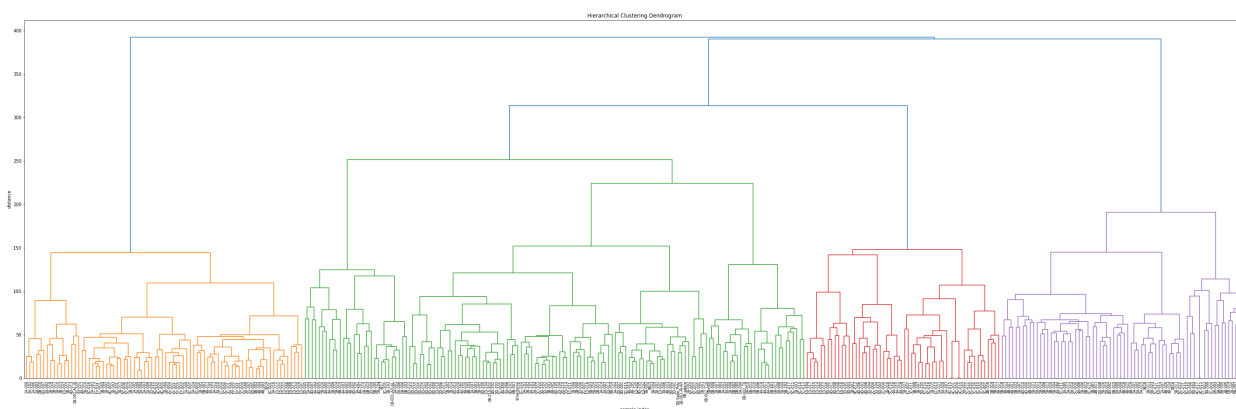


(b) Number of clusters along hierarchy

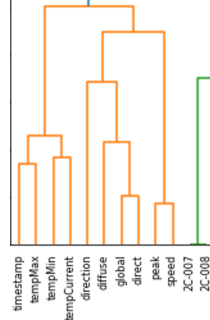
**Figure 3: Temperature sensors clustering**



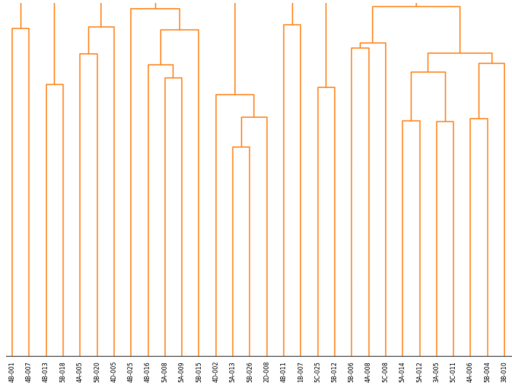
**Figure 4:** CO2 sensors clustering with weather data



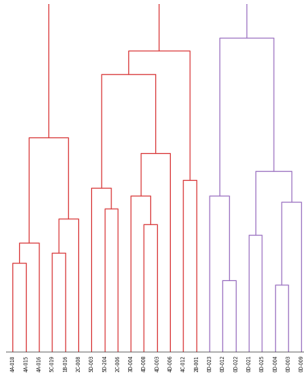
**Figure 5:** Temperature sensors clustering with weather data



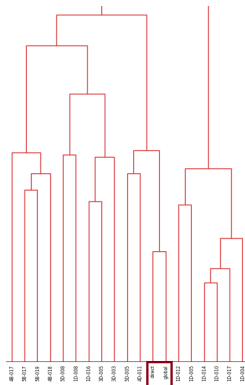
**Figure 6:** Zoom on weather features clusters



**Figure 7:** Zoom on clusters - with temperature



**Figure 8:** Zoom on clusters - with temperature



**Figure 9:** Zoom on direct and global sunlight features in clustering - with temperature

not utilize the workload characterization to solve a concrete problem. We only provide some guidelines for choosing appropriate application contexts where our data could be used in the future.

In (14) authors performed an extensive study of medical applications that could be executed on the Internet of Medical Things (IoMT) from an edge computing perspective. They performed a workload characterization of potential IoMT applications and explored the micro-architecture implications of these applications. They compared the workloads' characteristics to an existing embedded systems benchmark suite to reveal their differences and similarities. The critical challenge of the paper is to understand the computational capabilities and microprocessor architectures that satisfy IoMT applications' execution requirements while adhering to the devices' resource constraints. As one preliminary step for this intellectual challenge, one first needs to understand the execution characteristics of the applications that will potentially execute on IoMT devices. Contrary to this work, ours does not deal with applications producing data, and we only guess that individual sensors make independent data that we collect for analysis.

In (20) authors focused on Distributed Stream Processing Systems (DSPs) are hosted in Cloud data centers. DSPs are vital for real-time data processing and analytics in any IoT software architecture. In this context, the authors developed RIoTBench, a real-time IoT benchmark suite, along with performance metrics, to evaluate DSPS for streaming IoT applications. One key point of the paper is that the applications are based on common IoT patterns for data pre-processing, statistical summarization, and predictive analytics. Then, patterns are coupled with four stream workloads sourced from IoT observations on smart cities. Finally, the authors validated the RIoTBench suite for the widespread Apache Storm DSPS on the Microsoft Azure public Cloud. Contrary to this, our work questions the nature of data, and we do not focus on the applications producing streaming data.

In (19) authors focused on a detailed characterization of the servIoTicy platform, state-of-the-art infrastructure for hosting Internet of Things (IoT) workloads in the Cloud. ServIoTicy provides multi-tenant data stream processing capabilities, a REST API, data analytics, advanced queries, and multi-protocol support in a combination of advanced data-centric technologies. In short, the paper focused on the characterization of a system for IoT, not on IoT data sources. This paper is unaware of the nature of the data, contrary to our work. The system is supposed to serve any sources. If people want a dedicated service to



merge and analyze data, it should be integrated into the data processing pipe, a dedicated component of the whole architecture. Contrary to this work, ours does not focus on a system but on classifying and analyzing IoT data sources to extract knowledge.

Regarding the problem of data fusion, papers (2, 3) are seminal papers. Both of them proposed a rigorous axiomatization of the data fusion process. Contrary to these important works, we decided not to investigate the axiomatization issues because they are not the core of our concerns. We prefer to do a fusion on our data sources in a "natural" way, meaning that we do not automate the fusion process since the number of sources is kept low.

In (16) authors studied the electricity consumption in the Greener data set, a data set related to a tertiary building in Grenoble, France. Authors pushed the idea that energy disaggregation methods, i.e., the Non-Intrusive Load Monitoring (NILM) methods (10), are powerful tools for feedback on energy consumption. We also use the Greener data set but not the one related to the electricity consumption. We focus on the global quality attributes of the building and the CO2 metrics produced on multiple sensors located in various rooms. We provide orthogonal and complementary views of the building. Authors in (16) only consider the data quality aspect of the evaluated data set.

In (18) authors focused on using advanced artificial intelligence (AI) methods to optimize building energy usage while maintaining occupant thermal comfort. One key point is the definition of "comfort." The authors reviewed some known definitions and noticed that the question is still controversial. In our work, we do not go in the direction of thermal comfort because some data that come with the definition are unavailable.

In (17) authors noticed that Lack of spatiotemporal awareness could lead to excessive sensors or non-optimal distribution of sensors across a building. We extend the optimal sensor placement problem to find the minimal sensor group that can robustly approximate the rest of the sensor values. The solution uses a distributed learning step to parallelize the predictive capability of candidate groups generated by evolutionary computing techniques. They experiment on 24 zones from a seven-floor building in Thailand to establish a trade-off between virtualization accuracy, energy footprint, and installation cost. The work exploits the concept of "less is more" to bring down the capital and recurring expense for a smart building solution.

In (18) authors focused on using advanced artificial intelligence (AI) methods to optimize build-

ing energy usage while maintaining occupant thermal comfort. One key point is the definition of "comfort." The authors reviewed some known definitions and noticed that the question is still controversial. In our work, we do not go in the direction of thermal comfort because some data that come with the definition are unavailable.

## 5. Conclusion

Experiments with massive deployment of smart sensors are becoming increasingly numerous, not only in research laboratories but also in cities. The COVID pandemic, for example, pushed the deployment of CO2 sensors in schools. Even if it is regrettable, this is a fact that a sober deployment and a reasonable a priori on the use of the data produced and the objective to be reached have not always been used to guide the deployments.

Our work proposes a series of tests to characterize, before deployment, the data produced at the building level. The issues that we feel are important to address are related to the nature of the data, for example, the absence or presence of outliers and their management as they may impact the clustering, and thus the nature of the data models resulting from the analysis.

We propose in the article to build different views of the data. This process is a step forward so that the various stakeholders working on designing intelligent buildings can now rely on new business knowledge. We have proposed different analyses and observations on real data to illustrate the approach. We also discussed the conclusions we can make from the observations. Some of them are general, others less so, but all call us to question ourselves more on what should be a deployment that simultaneously considers data exploitation and exploration. Finally, based again on our actual data and as an illustration of our concern, we have proposed research issues to go further in the analyses, and their crossings, for reasoned and sober deployments of sensors from the IoT.

## References

1. S. J. e. a. Batz T Herzog R, "IoT Workload Emulation for Data Centers", tech. rep. (Open Research Europe, 1:12, [version 1; peer review: 1 approved]., 2021).
2. S. Benbernou, X. Huang, M. Ouziri, *IEEE Trans. Big Data* **7**, 436–450, (<https://doi.org/10.1109/TBDATA.2017.2710346>) (2021).

3. S. Benbernou, M. Ouziri, presented at the 2017 IEEE International Conference on Big Data (IEEE BigData 2017), Boston, MA, USA, December 11-14, 2017, ed. by J. Nie *et al.*, pp. 74–79, (<https://doi.org/10.1109/BigData.2017.8257913>).
4. B. E. Boser, I. M. Guyon, V. N. Vapnik, presented at the Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory, pp. 144–152.
5. M. M. Breunig, H.-P. Kriegel, R. T. Ng, J. Sander, presented at the ACM sigmod record, vol. 29, pp. 93–104.
6. S. Chandrasekaran, R. Subramaniam, *IEEE Spectrum* **January 11**, (<https://spectrum.ieee.org/why-iot-sensors-need-standards>) (2022).
7. J. M. Chenu, Ed., *Construire plus vite la ville durable - Nouveau modèle et partenariats* (Éditions Eyrolles, 2021), ISBN: 978-2-416-00257-1, (<http://www.crcpress.com/product/isbn/9781466558212>).
8. M. Ester, H.-P. Kriegel, J. Sander, X. Xu, presented at the KDD, pp. 226–231.
9. L. Goasduff, *Gartner Newsroom -Press Releases* **August 29**, (<https://www.gartner.com/en/newsroom/press-releases/2019-08-29-gartner-says-5-8-billion-enterprise-and-automotive-iot>) (2019).
10. G. Hart, *Proceedings of the IEEE* **80**, 1870–1891 (1992).
11. P. Horn, *IBM, Autonomic Computing: IBM's Perspective on the State of Information Technology*, ([https://people.scs.carleton.ca/~soma/bioseca/readings/autonomic\\_computing.pdf](https://people.scs.carleton.ca/~soma/bioseca/readings/autonomic_computing.pdf)).
12. J. O. Kephart, D. M. Chess, *Computer* **36**, 41–50, ISSN: 0018-9162, (<https://doi.org/10.1109/MC.2003.1160055>) (Jan. 2003).
13. M. Koziol, *IEEE Spectrum* **August 12**, (<https://spectrum.ieee.org/satellites-great-option-iot>) (2021).
14. A. Limaye, T. Adegbiya, presented at the 2017 IEEE Computer Society Annual Symposium on VLSI, ISVLSI 2017, Bochum, Germany, July 3-5, 2017, pp. 302–307, (<https://doi.org/10.1109/ISVLSI.2017.60>).
15. F. T. Liu, K. M. Ting, Z.-H. Zhou, presented at the 2008 Eighth IEEE International Conference on Data Mining, pp. 413–422.
16. G. F. Martin Nascimento, F. Wurtz, B. Delinchant, P. Kuo-Peng, N. J. Batistela, “GreEn-ER Data - Jeux de données de consommation d’électricité dans le secteur tertiaire”, working paper or preprint, Aug. 2021, (<https://hal.archives-ouvertes.fr/hal-03322581>).
17. A. Mitra, D. Trystram, presented at the.
18. J. Ngarambe, G. Y. Yun, M. Santamouris, *Energy and Buildings* **211**, 109807, ISSN: 0378-7788, (<https://www.sciencedirect.com/science/article/pii/S0378778819336527>) (2020).
19. J. L. Pérez, D. Carrera, presented at the 2015 IEEE First International Conference on Big Data Computing Service and Applications, pp. 62–71.
20. A. Shukla, S. Chaturvedi, Y. Simmhan, *Concurr. Comput. Pract. Exp.* **29**, (<https://doi.org/10.1002/cpe.4257>) (2017).
21. U. Tadakamalla, D. A. Menascé, presented at the Edge Computing - EDGE 2019 - Third International Conference, Held as Part of the Services Conference Federation, SCF 2019, San Diego, CA, USA, June 25-30, 2019, Proceedings, ed. by T. Zhang, J. Wei, L. Zhang, vol. 11520, pp. 1–15, ([https://doi.org/10.1007/978-3-030-23374-7%5C\\_1](https://doi.org/10.1007/978-3-030-23374-7%5C_1)).