Generation of explanations for energy management in buildings

Amr Alzouhri Alyafi

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : Informatique

Arrêté ministériel : 25 mai 2016

Présentée par

Amr ALZOUHRI ALYAFI

Thèse dirigée par **Patrick REIGNIER**, Professeur des Universités, Grenoble INP
et codirigée par **Stéphane PLOIX**, enseignant chercheur, G-SCOP

préparée au sein du **Laboratoire des Sciences pour la Conception, l'Optimisation et la Production de Grenoble** dans **l'École Doctorale Mathématiques, Sciences et technologies de l'information, Informatique**

Génération d'explications pour la gestion énergétiques dans les bâtiments

Generation of explanations for energy management in buildings

Thèse soutenue publiquement le **27 mai 2019**,
devant le jury composé de :

Monsieur PATRICK REIGNIER
PROFESSEUR, GRENOBLE INP, Directeur de thèse
Madame SALIMA HASSAS
PROFESSEUR, UNIVERSITE LYON 1, Rapporteur
Monsieur STEPHANE PLOIX
PROFESSEUR, GRENOBLE INP, Co-directeur de thèse
Monsieur JACKY MONTMAIN
PROFESSEUR, IMT MINES D'ALES, Rapporteur
Monsieur FRANÇOIS BREMOND
DIRECTEUR DE RECHERCHE, INRIA CENTRE S. ANTIPOLIS - MEDITERRANEE, Examinateur
Monsieur BRUNO PEUPORTIER
RESPONSABLE SCIENTIFIQUE, MINES PARISTECH, Examinateur
Monsieur ETIENNE WURTZ
DIRECTEUR DE RECHERCHE, INES/CEA - BOURGET-DU-LAC, Président

Contents

1 Smart Homes **9**

1.1 Introduction 9

1.2 Smart Homes 13

1.3 Persuasive interactive systems 17

1.3.1 FBM and TTM: two reference models for persuasive interactive systems 18

1.3.2 Persuasive design principles 19

1.4 Conclusion 21

2 Why explanations are crucial for occupants and what are energy services ? **23**

2.1 Knowledge and cooperation between occupants and the energy management system 24

2.2 Energy management systems 26

2.2.1 Smart thermostat 26

2.2.2 Rule-based energy systems 27

2.2.3 Model-based EMSs 29

2.3 Need for explanations in energy management systems 32

2.4 Explanations 34

2.5 Different forms of explanations 35

2.6 From sensory data to causality 38

2.6.1 Problem statement 39
2.6.2 Building a causal model from a bottom-up approach . 40
2.6.3 Injecting knowledge in the building process 41
2.7 Quality of an explanation 42
2.8 Conclusion 43

3 How to generate explanations 45
3.1 Qualitative knowledge 47
3.2 Knowledge injection 55
3.3 Causal graphs 58
3.4 Bayesian networks 58
3.5 Conclusion 60

4 Explanations generation with knowledge models 63
4.1 Problem statement and general solving principles 66
4.2 Description of experimental testbed 67
4.3 Physical knowledge model for building performance simulation 70
4.3.1 Thermal model 70
4.3.2 CO_2 based air quality model 72
4.3.3 Energy cost model 72
4.3.4 Encoding the solution vector 73
4.4 Generating explanations 74
4.4.1 Differential explanations 77
4.4.2 Differential explanations with contextual causality . . . 83
4.4.3 Model fragment 85
4.5 Direct explanations 88
4.6 Conclusion 93

5 Explanation generation without knowledge models 95
5.1 Contextual statement 96

5.1.1 The challenge of extracting similar days 97
5.1.2 The challenge of variables weighting 98
5.2 Proposed solution 99
5.3 Normalizing data in the building 100
5.4 Similarity method between different days 102
5.5 Extract similar days based on Context 107
5.6 Extract the best set of actions 110
5.7 Dataset 111
5.8 Validation and results 112
5.8.1 Knowledge model for validation 112
5.8.2 Validating the data model 113
5.8.3 Validating the recommended set of actions 116
5.9 Analysis and challenges 119
5.10 Generating differential explanations from the data model . . . 121
5.11 Generate Direct explanations from the data model 122
5.12 Conclusion 124

6 First user evaluation 125
6.1 Text Generation 125
6.1.1 Objective 126
6.1.2 Problem statement 127
6.1.3 Ariane-Heloise system 127
6.1.4 GRA-FRA 128
6.2 Human-machine interface 130
6.2.1 User interface requirements 130
6.2.2 Designed for daily life: catching the eye at the right time 132
6.2.3 User interface in detail 133
6.3 Validation scenario for the generated explanations 136
6.3.1 Context and goals 136

6.3.2 Method 137
6.3.3 Participants 138
6.3.4 Independent variables 138
6.3.5 Tasks 138
6.3.6 Scenario of the interview 139
6.3.7 Results 144
6.4 Second Validation 145
6.5 Conclusion 148

7 General conclusion 151

8 List of publications 155

9 Appendix 157
9.1 Optimization criteria 157
9.1.1 Compromise of interest 160
Automated decision making (the default paradigm) . . 161
Decision making by user intervention (the optional paradigm) 162
9.1.2 Optimization results and discussions 163
Choosing a many-objective optimization algorithm . . 164
Results of a random day for analyzing the efficacy of each objective 169
Year-round results to study seasonal variations 171

List of Tables

4.1 Description of the symbols used 64

5.1 Table of features in the building 112

9.1 Comparison of optimizer performance 167

Chapter 1

Smart Homes

1.1 Introduction

Since the beginning of the industrial revolution until our day, humans are demanding more and more energy. Each year there is an increase in the demand for energy all over the world. The American energy information administration [35] projected a rise by 28% of the global energy demand from 2015 to 2040 as shown in the Figure 1.1.

To be able to satisfy those demands, governments need to construct substantial electrical generation plants and adapt the network for the peak hours of demand. This is a considerable cost for countries. There is also another indirect cost of the same importance on the environment, as energy production is responsible for 25% of the emissions of greenhouse gases (GHG) and pollution on earth [1].

At the same time, there is a lot of wasted energy on the consumer side and especially in buildings, because buildings, in general, are the primary consumers of produced energy. According to statistics from the French Ministry of the Environment, Energy and Sea published in 2017 [26], 45% of the total energy consumption in France can be attributed to the residential

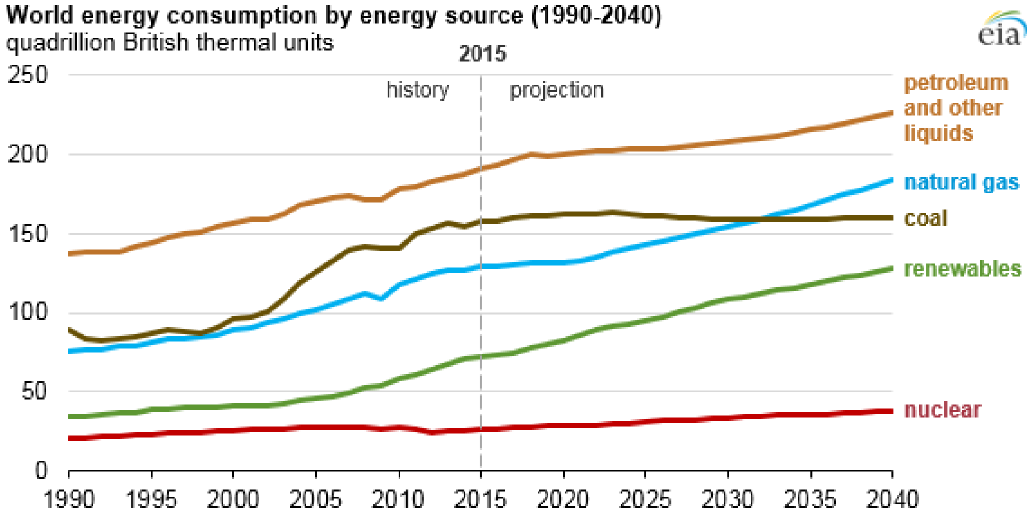


Figure 1.1: The American energy information administration (IEO2017)[35] projects that world energy consumption will grow by 28% between 2015 and 2040.

and tertiary sectors as shown in Figure 1.2. For example, France has more than 31 million residential buildings covering an area of more than two billion square meters. Commercial buildings account for more than 900 million square meters. Twenty million dwellings had been built before the first thermal regulations were introduced in 1975. Highly demanding of energy, these dwellings represent 58% of the housing sector and account for more than 75% of its energy consumption. Therefore, optimizing energy performance of the buildings is becoming extremely important.

Critically, stopping the waste of energy in dwellings is essential to fight the extra demand on power and minimize the greenhouse gas emissions. Therefore in the Paris agreement in 2016 to combat climate change, one of the targets was to try to reduce the consumption by 20% from 2016 to 2020 [20].

All of this also helps the inhabitants in maximizing their comfort and security while reducing waste, lowering energy bills, and helping them to

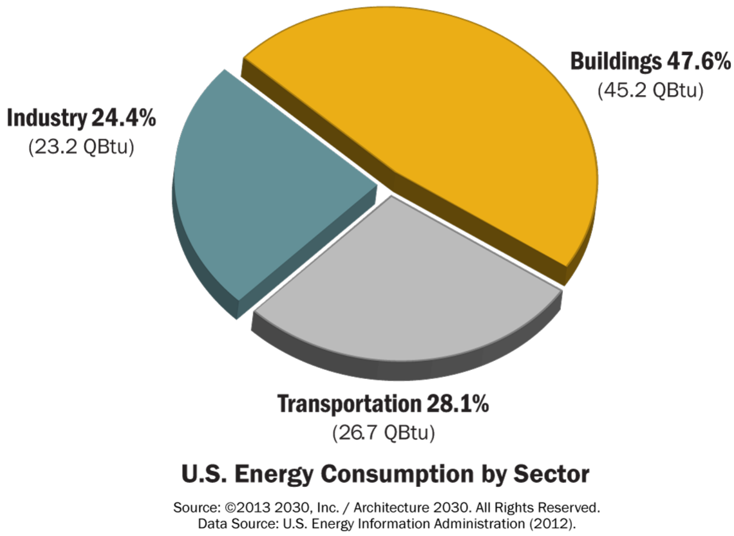


Figure 1.2: Buildings contribute more than 40% of global energy production

detect recurring problems [48].

Reducing energy waste and helping occupants were the main reasons for developing energy management systems (EMSs). An EMS is a system of computer-aided tools used by occupants to monitor, control, and optimize the performance of their energy consumption/production. Usually they are built using one or more physical models to simulate and optimize the energy usage. EMSs are very beneficial for occupants [106], but at the same time, constructing self-tuned energy models is still an ongoing scientific problem[105] [5], as they require profound physical knowledge to be built. They also need to be adapted for each thermal zone, and to be re-tuned or reconstructed with any change in the environment (like adding a mobile electrical heater). EMSs are dependent on sensors for data acquisition. Therefor EMSs add an extra cost to the sensors and their installation (fixing, cables,...) for occupants. This extra cost was the main reason that EMSs were initially only installed in some industrial and commercial buildings.

Sensors now have their own source of energy (like the EnOcean sensors that uses energy harvesting technologies)[115] and thanks to their wireless protocols, they can be easily deployed in an existing environment and for different applications like those shown in Figure 1.3.

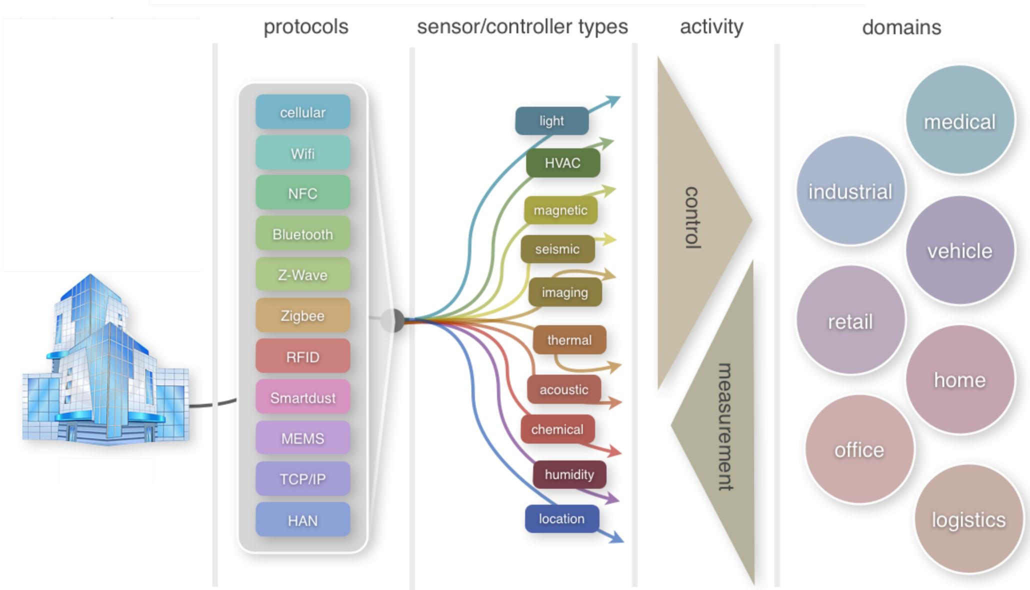


Figure 1.3: The concept of the Internet of Things

The drop in sensor prices and the annually excessive need for energy pushed authorities to change construction regulations. Authorities have introduced laws mandating the installation of different sensors, like fire detecting sensors for security, and smart meters to measure real consumption (like the new law in France RT-2012[22]). Further, governments have added different regulations for minimum building performance, and hourly pricing for electricity to fight the waste of energy, especially in the peak hours [22]. All of these technological and regulations achievements opened a new market for companies and this led to the creation of the smart homes.

1.2 Smart Homes

Major novelties were introduced into homes, principally because of new technological breakthroughs[103]. Changes cover such diverse things as the introduction of fixed fireplaces to the introduction of broadcasting, television and much more in the latter part of the 20th century. With the advancement in the Internet of Things (IOT) technology and the price drop for their devices, companies saw a huge market opening up and the population started hearing about smart homes. A smart home is defined as [56]: "A dwelling incorporating a communications network that connects the key electrical appliances and services, and allows them to be remotely controlled, monitored or accessed."

Companies detected an opportunity to sell new devices to allow remote control, automation and communication with other devices named "smart things". This market for smart homes and smart buildings is estimated to be one billion US dollars by 2020 [71] as shown in Figure1.4. Large companies like Google, Apple, Samsung, and others are now major players in this market.

Smart homes and automation

Traditional approaches to complex building management promote automation as a ***doing instead-of*** paradigm (i.e., delegation to an automated system). A mathematical model of the thermal characteristics of the house and its devices is employed. Then an optimization technique is applied to generate an optimal daily plan of actions according to a user's objectives. The plan is used by the energy manager to control the home. From the end user point of view, the house acts autonomously. Also, end users might not understand the system behavior and even feel that they have to fight against their home to reach their objectives. A typical example of this situation is a user who closes the blinds to adapt light conditions for watching TV. As a result, the system

Figure 1.4: Smart homes market share [71]

automatically switches the lights on because the room is too dark, which might be inappropriate. Moreover, a couple of minutes later, the control system may re-open the blinds to heat the room on a sunny winter day; although the equations that model the home and the objective functions appear to be complete from the point of view of physics, they do not take into account human factors such as intentions. As a result, the system has good properties from the theoretical perspective of energy consumption but may be inappropriate for the inhabitant's daily life.

It is clear that the ***doing instead-of*** paradigm might not meet the occupant's goals because the system will not be able to understand the user's objectives nor his needs. Also from a psychological point of view, inhabitants like to feel in control [8]. For example occupants prefer to open the window by themselves and not by actuators, because this reassures them, give them the feeling of being in control and that needs to be taken into consideration as a part of their comfort [8].

An energy manager has a partial knowledge of the environment but poor knowledge of the inhabitant's intentions or profile. The inhabitant knows what he wants and which comfort level is important for him but he has poor knowledge about how to reach that. In conclusion, the user and the energy manager have to cooperate to lower the energy consumption while maintaining an acceptable level of comfort .

This indicates the need for a method to cooperate and interact between occupants and EMSs, to give them the ability to understand how the EMS is working and making decisions.

The INVOLVED ANR project, which this thesis is part of, aims to build a persuasive interactive system. This research project explores a new paradigm: an energy advisor (e-consultant) who can interact and cooperate with people to gather information, suggest actions, show and explain. In particular, this advisor must be able to create a type of cooperation between inhabitants and energy systems to help occupants understand the energy functioning of their habitat and to assist them in achieving their objectives.

Figure 1.5, presents the different difficulties that occupants face regularly (on the left) and the different services proposed by the INVOLVED project (on the right) to answer those difficulties.

Different services are identified by INVOLVED to help occupants understand their energy system, and push them to start adjusting their routines and take into account energy in a better way. Seven services are presumed to be useful for occupants to increase their awareness about their energy consumption and their energy system:

- MIRROR: reflect user behavior.
- WHAT-IF (Replay): determine by simulation the consequences of a

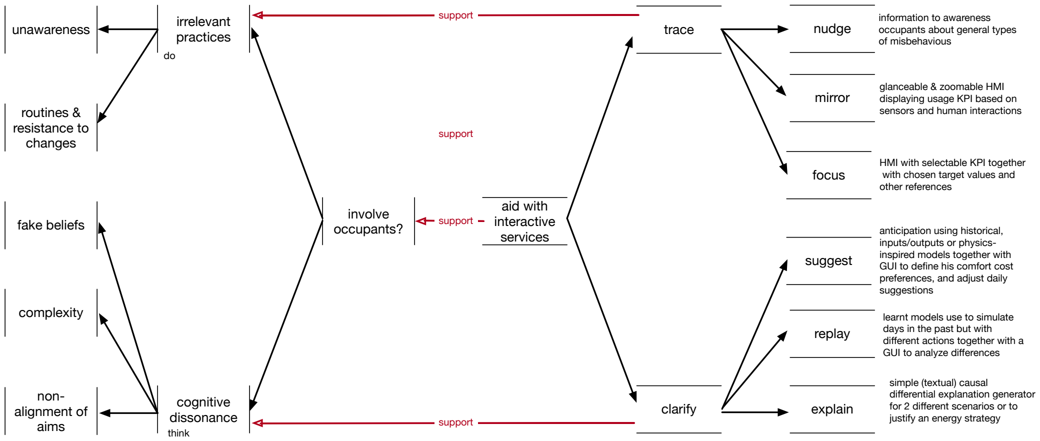


Figure 1.5: Involved interactive services for occupants

behavior.

- EXPLAIN: explain to users the manager's decisions and calculation results to improve their understanding of the phenomena at stake in their home.
- SUGGEST-AND-ADJUST: calculates an optimal consumption strategy for the next period
- RECOMMEND: diagnose past behaviors to advise new actions or strategies to reduce consumption.
- FOCUS: Human-machine interface(HMI) display usage key performance indicators (KPIs) based on sensors and human interactions. It also allow the user to zoom in on his objectives.
- NUDGE: send information to awareness occupants about general types of misbehaviors

This thesis focuses on the EXPLAIN feature. The EXPLAIN feature allows the user to ask the EMS about the consequences of his own decisions. These consequences are estimated, for example, by simulation from physical models. The result of a query is the final state obtained but also the explanation describing why he is getting those consequences. It is, therefore, a question of allowing the end user to easily specify an initial situation (for example, depending on historical measurements of sensors) as well as the evolution scenario he wishes to test.

Explanations should allow the user to estimate the impact of his actions in a complex environment, which means that explanations should present the impact of user actions and its alignment with the user objectives.

The INVOLVED goal is to build a persuasive system to sensitize occupants about their energy consumption and persuade them to adopt better behavior from an energy point of view. The next section describes the principles of this persuasive technology and how it can help to change the occupant's behavior.

1.3 Persuasive interactive systems

Persuasive technology refers to "an interactive technology [which] aims at changing a person's attitudes or behaviors" [39].

Persuasive technology is defined as a technology that is designed to change attitudes or behaviors of the users through persuasion and social influence. These persuasive systems are expected to sense inappropriate or undesirable behavior such as over smoking[70], and then enact functions, such as recommendations, suitable to support users' changes. The field draws upon a large body of models and theories such as the technology acceptance model [24], and the Motivation – Opportunity – Ability model [77]. Two seminal models that currently serve as references for the development of interactive persuasive systems are: the Fogg Behavior Model (FBM) and the Transtheoretical Model of change (TTM). The next section summarizes the design principles drawn from these models and illustrates the discussion with representative examples of persuasive interactive systems related to energy management and water usage.

1.3.1 FBM and TTM: two reference models for persuasive interactive systems

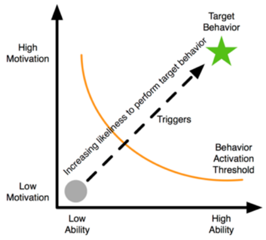


Figure 1.6: Fogg's Behavior model

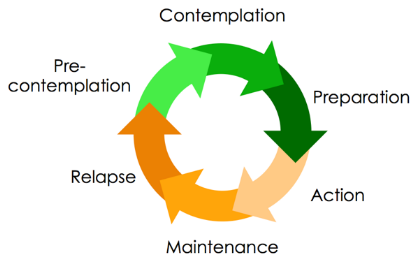


Figure 1.7: Transtheoretical model of change

Fogg's Behavior Model characterizes human behavior according to three dimensions [38]: motivation to adopt a particular target behavior, ability to perform the target behavior, and triggers. As shown in Figure 1.6a, motivation, and ability define a 2D-space where both human behavior and resistance to change can be characterized. Any trigger received by an individual when being above the activation threshold in the 2D-space should make this person adopt the behavior. In turn, motivation can evolve along three dimensions: pleasure/pain, hope/fear, and social acceptance/rejection. Learning new skills being not frequently accepted by people.

The TTM [96] [95] decomposes behavior change into a 6-stage cycle (Figure 1.7):

- Pre-contemplation: subjects are not considering the idea of change, maybe because they are unaware or not informed or possibly frustrated by a previous failed change attempt. They do not intend to take action.

- Contemplation: subjects are aware that they should change a certain behavior, and they consider attempting the change. In this stage, they try to get informed about the problem, but they are not ready to take concrete action for changing.

- Preparation: subjects are ready to make a change in the near future (usually measured as the next month), they are trying to develop a plan to take their first concrete action in the direction of the change.

- Action: subjects are moved to action and have modified their behavior.

- Maintenance: subjects try to keep the behavior change, and struggle to prevent relapsing. If they fail at this stage, relapse will occur, regress them to an earlier stage, and they will have to restart the progress from the first stages.

- Relapse: subjects fail at the target behavioral change.

This TTM describes the normal inhabitant who has the objective of enhancing his energy consumption.

1.3.2 Persuasive design principles

According to Fogg, technology can play three roles [40]: as a tool, as media, and as a social actor. As a tool, technology can make activities easier or more efficient to perform. The corresponding design principles are then 're-duction', 'tunnelling', 'tailoring', 'suggestion', 'monitoring', 'self-monitoring', and 'conditioning'. In particular:

- Suggestion: persuasion power can be increased by offering a suggestion about behavior change. One of the key features of the persuasive system is its capacity to provide inhabitants with recommended action plans along with the appropriate contextual explanations;

- Monitoring and self-monitoring: technology "eliminates the tedium of tracking performance". As such, it serves as the basis for revealing behavior and for monitoring progress. Mirroring inhabitants behavior and their impact on comfort level and energy savings is part of the e-consultant solution;

- Conditioning: positive reinforcement can be used to "transform existing behaviors into habits".

As media, technology can shape attitudes and behaviors by providing compelling simulated experiences. The corresponding design principles are 'cause and effect', 'virtual rehearsal', 'virtual rewards' and 'simulations in real world contexts'. For instance, Fogg [40] defines the 'cause and effect' principle as a means to persuade people to change as a simulation can make observable "the link between cause and effect". The next chapter will show how this "cause and effect" principle is applied to the generation of contextual explanations, as well as the "what-if" feature that allows inhabitants to simulate the effects of alternative behavior on energy consumption.

As a social actor, technology persuades by giving a variety of social cues that elicit social responses from their human users. The corresponding design principles are 'attractiveness', 'similarity', 'praise', 'reciprocity' and 'authority'. Among these principles, focusing on mobility, 'social comparison' is another but related principle [40]: performance comparison with the performance of others can increase motivation. In its current version, the social dimension has not been investigated for energy systems.

In an example, [21] the user will have to trace his own water consumption along with his activities. At the end of a period of time (a week for example), he can get feedback about his different activities like those shown in Figure 1.8. Then he can interact with the system and check the impact of each activity and what he can do to minimize it. In this example, it is possible for the end user to interact with the system. The system makes energy visible and allows the "what-if" interactions.

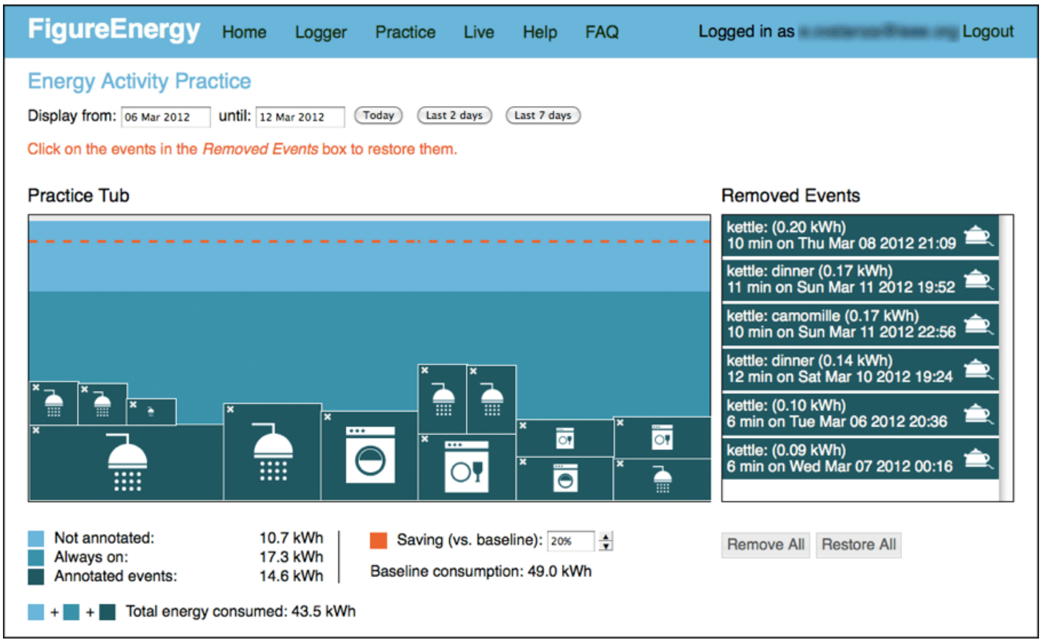


Figure 1.8: From user interface to user interaction in Costanza et al. [21]

This section presented persuasive technology and showed the importance of cooperation between the user and the system to achieve better results. It also illustrated how causal explanations can be useful in persuading occupants to change their behavior and adopt a new one.

1.4 Conclusion

Energy consumption in buildings is crucial in the energy market, and one of the key elements for reducing the waste of energy.

Due to the complexity of different phenomena and the various changing variables presented in homes, occupants are unable to correlate and understand how their habitat is functioning.

At the same time, occupants are capable of determining their comfort and their intentions. This shows that neither occupants nor the energy system can act alone. They both need to cooperate to achieve the occupant's objectives.

Explanations are a helpful tool for the cooperation between the EMS and occupants. Through them, occupants will be able to understand how their habitat is working, the impact of their actions, compare and challenge themselves with other people or with their own past results. They can also play an important role in persuading the occupants to change their behavior toward a new one with less energy waste.

The next chapter discusses the difficulties lying behind the generation of explanations and the scientific problems behind them. It explains at the same time, why it is so important to solve those problems so as to involve the occupants in the loop with the EMS.

Chapter 2

Why explanations are crucial for occupants and what are energy services ?

Energy is essential in our modern life but there is a lot of energy wastage in our homes. Occupants are disconnected from their energy system as they cannot easily understand both energy phenomena and the strategies proposed by their energy management system (EMS). However, the knowledge and information known by the EMS is different from that of the occupants. Like the example in the first chapter, the user wants to watch television with low light but the system estimates that the room is too dark and puts the lights on. Both of them have part of the total knowledge but they do not cooperate. Both the EMS and the occupants are important to obtain an acceptable comfort level and optimize the energy consumption as their knowledge is complementary; this will be expanded upon and clarified throughout the chapter. This chapter presents a brief look at the different existing energy management systems and their functionality. Then, it continues with explanations functionality and their utility. The chapter illustrates why explanations are useful for occupants, what is the most appropriate form of explanation to communicate to the occupants, and why it is difficult to generate those explanations.

2.1 Knowledge and cooperation between occupants and the energy management system

As stated in the first chapter, knowledge is distributed between the occupants and the EMS. Figure 2.1, illustrates the different levels of knowledge owned by each actor (Occupants, Expert, ...). It also represents how they interact and cooperate with the environment and with each other.

The energy system acquires the information about the environment through the different installed sensors, e.g. temperature, humidity, and light, sensors. That is why the information in the EMS is essentially quantitative but limited by the number and type of installed sensors. On the other hand, occupants and experts are another source of information for the EMS. They can provide the system with different observations (room size, window size, orientation, neighborhood,...), set points, and feedback for the system.

Occupants have their own, and complementary, information like comfort and intentions. Occupants' knowledge is more qualitative and up to date than the EMS's. Its level varies depending on each occupant.

To achieve an acceptable level of comfort without extra cost, neither the EMS nor the occupant can achieve that alone; they need to cooperate. This cooperation can be improved by explanations. Explanations describe for occupants why the energy system is recommending different actions at different times. They also provide the occupants with more insight into what they might obtain if they change their actions. This will be clarified later in the chapter, after detailing the different types of EMSs and their knowledge.

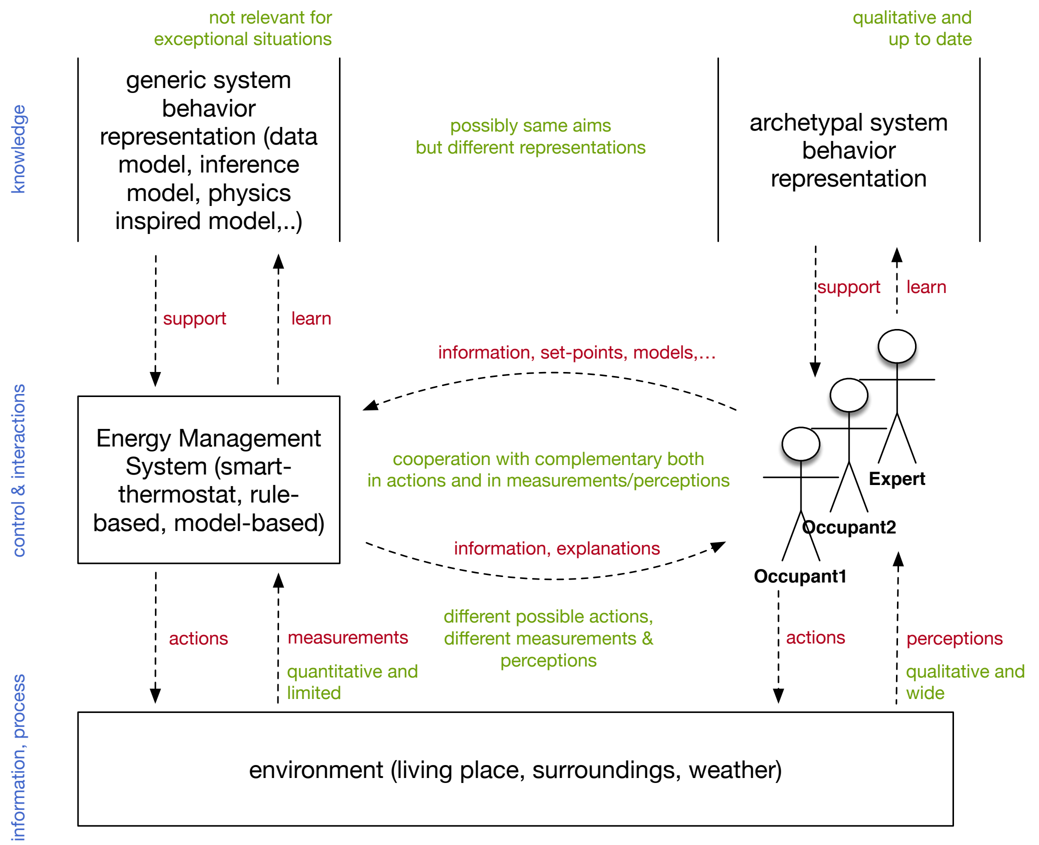


Figure 2.1: Knowledge and cooperation between user and the energy management system

2.2 Energy management systems

Energy management systems and human-machine cooperation are developing over time. They have different forms with different levels of complexity. Some of them are simple, like the smart thermostats, others are more complex, like the rule-based EMSs, or the model-based EMSs. Model-based energy systems have the capability to simulate or predict the evolution of the built environment with different changes.

2.2.1 Smart thermostat

Smart thermostats (like Nest, Tado, Ecobee, Netatmo... which are some of the existing brands in the market) are devices that can be used with home automation and are mostly responsible for controlling a home's heating and air conditioning. They perform the same functions as a programmable thermostat as well, as they allow the user to control the temperature of their home throughout the day using a predefined or learnt schedule, such as setting a different temperature at night.

Smart thermostats are usually connected to the Internet. They allow users to adjust heating settings from other Internet-connected devices, such as smart phones. In addition, they enable users to regulate the temperature remotely and with ease.

Smart thermostats also record internal and external temperatures, and how long the HVAC (heating, ventilation, and air conditioning) system has been running. This information is typically displayed later on an Internet-connected device. More intelligent thermostats appeared in the market, like the Nest thermostat, in Figure 2.2. These intelligent thermostats try to learn from the occupant's preferred temperature in different conditions, and also try to determine if occupants are wanting to save energy.

Smart thermostats remain a useful tool for occupants due to their ability

Figure 2.2: Nest smart thermostat from the nest labs

to help in reducing the waste of energy, especially with the HVAC systems[50]. Smart thermostats contain little knowledge about home contexts. They are not capable of predicting either the user's activities or judging whether those activities are going to result in a positive or negative impact on the occupants' comfort criteria. For example, if the occupant opens the window while the HVAC system is ON, the smart thermostat will not be able to detect the action or estimate if this is a good / bad action. Also, it will not be able to determine how much opening of the window at a certain time is impacting the comfort criteria.

In these type of systems, the cooperation and interaction between occupants and the smart thermostat is limited. The user can define only the set points and some states. The smart thermostat can only display the current value and it might give some general energy advice.

2.2.2 Rule-based energy systems

These systems are based on **IF-THEN** rules. These rules are usually used to build a typical case scenario or different ones to reduce the waste of energy in dwellings. Figure 2.3 presents a part of a rule-based model. Here the system will check if humans are detected [59]. Afterwards, it checks whether the light is ON or OFF. If it is OFF, the system will send a command to put the light ON.

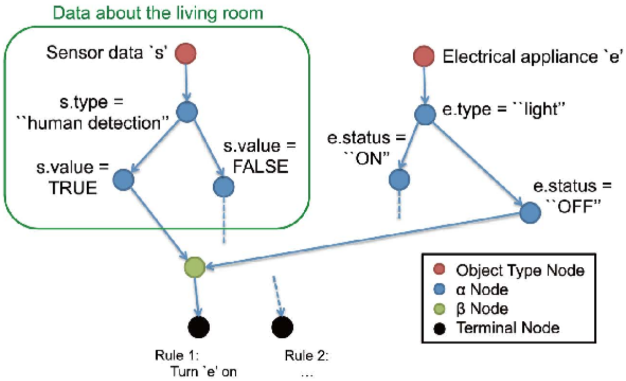


Figure 2.3: An example of a rule-based system [59]

However, when there are many appliances and many measurements from different sensors, this approach will be much more complicated to implement as there is a need to define how the system should react with different contexts and should cover various different scenarios of use.

Rule-based EMSs do not provide the ability to simulate or judge the evolution of the environment. For example, they do not have the capacity to predict how the occupant's comfort would change when the environment variables change.

In the rule-based EMSs, the user can interact more with the system than with the smart thermostats. The user can set different information about the context, the home state, and also personalize some scenarios. The system can display the home state, some indicators, and different general advice.

2.2.3 Model-based EMSs

Due to the complexity of buildings, and especially when trying to integrate the interactions between the occupants and the habitat, there is a need to create behavioral models of the living zones. This approach is usually studied in research and appears in many different studies[74][113]. Besides its complexity, this approach is the only one capable of tracking the evolution of the environment.

Energy models vary depending on the amount of knowledge injected into each one of them. They are categorized into three types:

1. **White box model:** Is completely built with physical knowledge, for the different aspects and diverse phenomena. Detailed physical knowledge models exist because they are compulsory for buildings design, but they do not fit well with reality although they can be tuned to match measurements. The expert building the model needs competence in physics and knowledge of all of the characteristics of the building (location, size, wall thickness, used materials,...), and the installed appliances, to be able to construct this kind of model.

2. **Black box model:** [105] Black box models are based on a general purpose structure (like linear regression models). This type of model is suitable for parameter estimation and does not require any expertise in physics. Although black box models are often criticized for their lack of physical interpretation [37], Richalet [98] demonstrated that certain kinds of these models allow the recovery of some physical information. Universal models can be either linear or non-linear and accept one or several inputs. Most of the time, universal models have been used for modelling specific systems or walls rather than the whole building [19], [114]. Extending these models for describing a whole building system

would be very costly in time and calculation power depending on the number of variables [105].

These models contain little physical awareness, so it is usual to obtain different results which contradict with the physics.

3. **Gray box model:** With gray box models, researchers mix data from the two models above to benefit from the knowledge in the white box that determines the structure of the model, and in parallel, use the algorithms in the black box model to simplify the formulae and accelerate the tuning phase of the white box[115].

   This is why gray box models are the most used outside of laboratories, as although they are less accurate than the white box models, where structure is still consistent with the physics, they are easier to build.

As identified, the central element of a model-based EMS is one or more energy models based on the characteristics of the various components of a building. These models allow the simulation and the prediction of the evolution of multiple phenomena (temperature, air quality, consumption, etc). They also enable the optimization process to compute the best energy management strategies according to some cost/comfort compromises.

A knowledge model generally speaking is a set of equations that describe housing, modelling the thermal properties of the walls, floors, and ceilings in terms of thermal resistance and thermal inertia, etc. These models are able to predict the evolution of physical variables (e.g. inside temperature, CO2 concentration, and humidity), by considering the values of these variables at the previous time step as well as the values of other environmental variables such as weather conditions, doors, and windows opening. Thermal models rely on the analogy with electrical circuits, like the example in Figure 2.4 while air quality models consist of differential equations. For instance, electrical resistance and capacitance represent thermal resistance and inertia. The electrical circuit is based on the characteristics of the housing. The CO2 model is a differential equation that uses room volumes, number of occupants and CO2 concentration in adjacent rooms. The various parameters of the thermal models are determined using genetic algorithms to fit the measurements of the sensors [106].

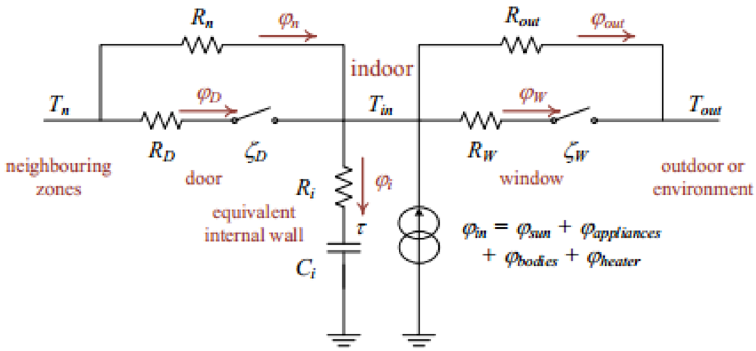


Figure 2.4: An example of a knowledge model [74]

Due to the model's mathematical formalism, the models themselves, and the optimization methods, are not suitable for direct communication with the occupants; the intrinsic knowledge they contain is not directly intelligible. Experts design energy models with the aim of reproducing the resulting phenomena without necessarily relying on the model's internal mechanism.

Still, an energy model (however complex it may be) does not possess all the necessary information to act correctly. The sensors give only a partial and noisy picture of the situation, and there will be information to which they will never have access such as the intent of the occupants or the addition of a mobile heater. The end user is the only one who knows what he wants to do, therefore he is the final decision maker, but he does not necessarily know how to do it correctly (with regard to energy expenditure) in relation to the constraints linked to the functioning of the building. He lacks the knowledge to make informed decisions. The next section will describe in more detail why explanations are useful for users.

2.3 Need for explanations in energy management systems

Explanations are useful in the persuasive technology, yet they have been also used with expert systems. The 80's have seen the rise of expert systems (and, more generally, knowledge-based systems (KBSs)) and their adoption in corporations as part of the information system. Expert systems are usually based on an explicit high-level symbolic knowledge database and an inference engine to exploit it. They are able to answer questions and take decisions in their expertise domains. Beyond the decision making aspect, the question of the generation of corresponding explanations was quickly raised and became an important field of study (see [47]–[91] for instance). Gregor et al. [47] stated that explanations in intelligent systems are important for several reasons. An expert user trusting his KBS will need explanations from the system mainly when he detects anomalies (when he will disagree with its indications). On the other hand, a novice user will use more the explanatory capacity of the system but with a short-term or long-term learning objective. By providing explanations, a system becomes more transparent. It will also appear to be more competent and trust in an automated system is related to a user's perception of its competence (see [82] for instance). Explanations are also required by the users when they lack some knowledge needed to contribute properly in a problem solving process. As mentioned before, neither the KBS nor the user can solve the problem alone so they have to cooperate. Users have their own expertise that may differ from the expertise of the KBS, and they know the context where the problem occurs. "Computational technology should be used, not to make or recommend solutions, but to help users in the process of reaching their decision" [58]. This is what Woods et al. called a "Joint Cognitive System" [118]. The system and the user are engaged in a cooperative process relying on explanations.

Karsenty and Brézillon have studied cooperative dialogs in natural working conditions [58]. Two different types of dialog were considered: validation dialogs and design dialogs. Validation dialogs are between an expert and a user. The expert is proposing a solution validated or modified by the user. The expert and user have complementary knowledge; the expert has the technical knowledge and the user has the domain expertise. In design dialogs, the first solution is not proposed by one of the experts but a preliminary solution is proposed from the problem definition by a close cooperation between both, each one bringing their technical expertise.

Context plays a central role in the man-system cooperation process when solving a problem. It is an important element in producing more pertinent explanations and thereby, enhancing cooperation. Defining context has been a challenging problem for the last 20 years (see [16] for instance for an early survey). Its definition may even depend on the context [17]. In the ubiquitous computing domain, Dey has proposed a definition focusing on man-machine interaction: "Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves" [29]. Context can be both considered from a representational or interactional point of view [17], [30]. Context as representation is information. It can be separated from activity. Context elements can be selected and described before the interaction occurs and are not modified by the interaction; activity happens within the context [30]. Context as interaction is a relation between activities or objects. It only exists during the interaction (task or problem solving) and evolves dynamically.

As mentioned in the first chapter, building a fully automated energy manager (doing instead-of paradigm) may be inappropriate for users. This work is positioned in the same situation as described by Woods et al.[118] or Karsenty et al.[58]: the system has some knowledge and competence to solve the problem, with the notable difference that this knowledge is numerical (physical knowledge, optimization capabilities etc.) and not symbolic and so is more intelligible, as in the studies cited above. The user has his own and complementary knowledge (like his own perception of comfort, his intentions etc). Neither the system nor the user can solve individually the whole problem but they have to cooperate (joint cognitive system). And this cooperation corresponds to what Karsenty et al. have described as a validation as opposed to a design dialog [58]: the expert has the technical knowledge, the user has the domain knowledge. The expert is proposing a solution that is accepted or modified by the user who makes the final decision. Cooperation is based on explanations. In a fully cooperative system, the energy system is modifying the user's perception of the problem but there is also reciprocity [58]. The user can modify the energy system by providing his own explanations to the system. The approaches presented in this work focus on the first aspect. The energy system is making propositions (energy plans) and provides explanations. This work has not considered yet the user feedback modifying the energy system's perception of its environment. The next sections quickly review explanations, and discuss the various forms of explanation and the quality of them.

2.4 Explanations

Explanations play different roles and take a variety of forms. Causality is also an important concept involved in the production of explanations for cooperation. These aspects are discussed in this chapter with a particular focus on the problem of elaborating a formal model of causality from sensory data. This chapter closes with a description of the qualities that explanations for cooperation should satisfy.

An explanation is a communication act between one or several people. Its main objective is to increase the comprehension for the receiver. As well as increasing knowledge, an explanation may have several other objectives, as stipulated in [60], (a survey article that has mainly inspired this section):

- To be able to predict similar events in the future [49].
- To be part of a diagnostic process [46]. Diagnosis is used to repair a malfunctioning system. It can also be used to reinforce an efficient or inefficient action to solve a particular problem. This association may help to inform the decision to choose, or not choose, the same action(s) to solve the same problem in the future.
- To justify an action [60].

All of these objectives, except the "diagnostic process for repair", are relevant to our goal.

Helping occupants to make optimal plans and demonstrating the efficiency of their actions will encourage them to act correctly in similar situations. Trust between an end user and an automatic system such as an EMS is similar to the trust between two people. Trust is related to the user's perception of the system competence [82], so an EMS that is able to justify its actions using explanations appears to be competent and will reinforce the occupant's trust in its correctness.

2.5 Different forms of explanations

As stated before, explanations are fundamental for occupants to understand and trust the EMS. Historically, explanations, as a scientific domain, were primarily studied in the social sciences (behavior explanations) and in philosophy (scientific explanations) [76]. For instance, Aristotle identified two types of scientific knowledge: the "knowledge about that" and the "knowledge about why". "Knowledge about that" is descriptive whereas "knowledge about why" is explanatory. "It is one thing to know that each planet periodically reverses the direction of its motion with respect to the background of fixed stars; it is quite a different matter to know why" [102]. Aristotle considered that a scientific explanation should be a list of deductive arguments. This early influential vision has led to the widely used Deductive-Nomological model (D-N) [51]. A D-N model of explanation is similar to a logical proof. It involves three types of statement: statements about the initial conditions, statements about laws and theories (both of them being called the *explanans*), and the observed statements that describe the phenomena to be explained (the *explanandum*). The *explanandum* is validly explained if it can be validly deduced from the *explanans* and the explanation is the deduction. In this process, all the statements must be true: the *explanandum* (which is obviously true because it is observed), the initial conditions and the laws. For instance, “all gases expand when heated under constant pressure” is a law. An initial condition can be “a bladder balloon is heated” and the observation is “the size of the bladder balloon is increasing”. The observation can be deduced from the initial condition and the law is the explanation. One of the limitations of this approach is the definition of the "law". A law should be a true generalization that does not suffer exception. Typically, a probabilistic phenomenon is not qualified as a law. The singular causal explanation of Michael Scriven [109] is also an example of a statement that will not meet the law criterion because it is not a generalization: "the impact of my knee on the desk caused the tipping over of the inkwell".

New models of explanations have been proposed to address the limitations of the D-N model. In particular, the logic formalism has been relaxed toward a variety of forms including textual representations, diagrams, and simulations. These forms may not capture the full complexity of a phenomenon, but they serve as useful approximations of a phenomenon. Here, causality plays a central role.

Causality is usually present in everyday life explanations. Philosophers of science have used this notion to illustrate why some things explain others and not the opposite. Causal relationships can be deterministic (like two pieces of a mechanism), or probabilistic (as in economy, psychology, etc.). However, explanations do not rely solely on causal relationships. For instance, one can explain how a mathematical result is achieved, or why China is bordered by 14 different countries [61]. Nevertheless, when both causal and non-causal elements are present in an explanation, the causal elements dominate the patterns of judgment [84]. There are at least four kinds of causal explanations:

- Common cause. The same cause has several effects, like a cold may cause fever and stuffy nose. It is often used in diagnosis.
- Common effect. Several causes converge to the same point. This is often the case in history. For instance, World War I was not caused only by the murder of Franz Ferdinand.
- Linear chain. There is a unique series of steps between an initial cause and a final effect. Linear chains are easy to understand but not very frequent in real life.
- Causal homeostasis. This is a causal relation that forms a stable cycle that reinforces itself.

Another way to differentiate between various forms of explanations (D-N, textual, causal, a-causal, etc.) is the use of stances [28]. A stance is one way to set the frame for an explanation. A stance can be mechanical, design or intentional. The mechanical stance considers that the phenomenon to be explained is composed of entities and is the result of their organization and interactions [10]. The explanation can be the mechanism itself (ontic) or the description of the mechanism (epistemic) [53]. The design stance considers that entities have functions and have been created on purpose. The explanation is based on that purpose and not on the mechanical functioning of the entity. Finally, in an intentional stance, the entity whose behavior is to be predicted is treated as a rational agent which, given its supposed beliefs and goals, will very likely act to reach these goals. Our primary objective is to help occupants to understand better how they can modify their behavior and how they can improve their decisions concerning their comfort criteria. Due to the complexity of the building's physics as well as to unconscious routines, occupants have difficulties in understanding what is happening and why they need to change their routines to improve their comfort, or how to make appropriate compromises between comfort and cost. Occupants do not need explanations as logical proofs (D-N), but as approximations that provide them with sufficient information about the current phenomena (epistemic mechanical stance). Explanations must not be generic but strictly related to their behavior as well as to the characteristics of their housing. As mentioned above, causality plays a central role. Generation of contextual and causal explanations is therefore vital.

2.6 From sensory data to causality

End user's housing is more and more equipped with sensors that measure environmental variables such as temperature, humidity, CO2 level, weather conditions, and the number of people in a room. Sensors also measure the user's actions such as turning on the heater or opening doors and windows. In our case, generating an explanation for cooperation requires the analysis of the data flow provided by the sensors as well as the identification of the causal relationships between the actions of the occupants and variation in the environmental variables. The next part will consider that a phenomenon is a value or the variation of the value of a variable. Some of the phenomena might be characterized as causes, some others as effects of these causes.

2.6.1 Problem statement

Causality from sensory data is difficult to model mathematically. Effects can be directly observed, but causal relationships cannot. Considering phenomena as events, a cause (C) always precedes the observation of an effect (E) but an effect (E) observed after (C) and correlated with it, does not necessarily mean that (C) is the cause of (E). The "car allergic to vanilla ice cream" scenario illustrates this case [63]: a man used to buy ice cream after dinner for his family. He complained to General Motors that every time he bought vanilla ice cream, he had difficulties in starting the car engine (other ice cream flavors were fine). General Motors engineers finally found that the cause of the problem was vapor lock. Actually, it took less time to buy vanilla ice cream than for other flavors. As a result, the engine remained too hot for the vapor lock to dissipate. The co-occurrence of buying vanilla ice cream and the car not starting did not mean that buying vanilla ice cream was the cause of the car failure.

As illustrated in Figure 2.5, the co-occurrence (with a potential time delay dt) of two phenomena calls for several interpretations: precedence only (2.5.a), direct causal relationship (2.5.b), consequences of a third phenomenon that may be outside of perception (2.5.c). For instance, having a flue may first cause fever and then coughing. Ignoring the existence of viruses may lead to the belief that the fever is the cause of coughing.

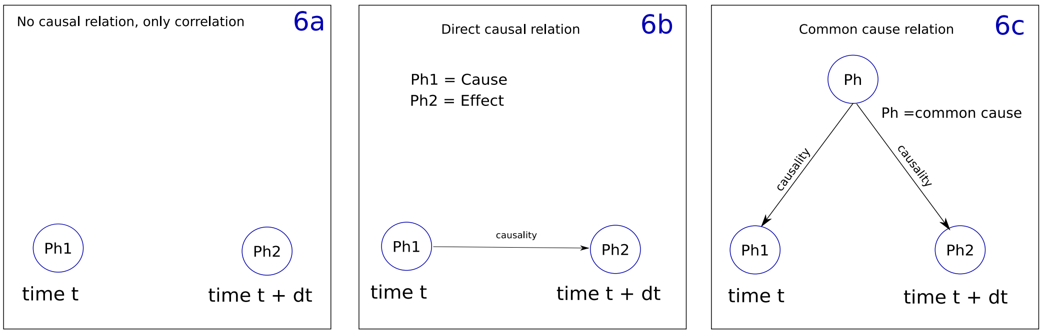


Figure 2.5: Causal relationships between co-occurrent phenomena

2.6.2 Building a causal model from a bottom-up approach

Extracting a causal model solely from time-correlated observations is a promising approach to the elaboration of a genuinely adaptive system. As almost no fundamental expert knowledge is necessary (everything is extracted from the data flow), such a system can be deployed in new housing and can learn a model of the environment from its interactions and observations. Constructivist or Developmental Artificial Intelligence is an example of such an approach. It has been originally applied to robotics, but there are some examples in the context of ambient intelligence (see [78] or [86] for instance). However, learning a causal model from a purely bottom-up approach is a complex task. In particular, one cannot be sure that all the critical phenomena are observed in order to infer “common cause” causal relationships. For a common causal relationship (Figure 1.5c), is it necessary to find the common cause? If this common cause cannot be determined or observed, we might conclude there is a direct temporal relationship between the two phenomena (as for Ph1 and Ph2 in Figure 1.5b). Direct relations are useful for predictions. If Ph1 is observed, it is known that Ph2 will then be observed. But because one has a partial view of the world, one cannot predict when this relationship might fail. In addition, since the model is based on observations but is not intrinsic to a mechanism, it cannot be used to determine which external intervention will modify its behavior. For instance, it can be observed that every morning, the rooster wakes up, sings and soon after, the sun rises. There is, of course, no direct causal relationship between these phenomena but it can be used to predict sun rising. Even so, there is no point in waking up the rooster earlier to get a longer day.

2.6.3 Injecting knowledge in the building process

Complexity arises from the size of the possibility space (all the possible causal models to explain the observed phenomena). One open question is 'How is our brain able to build complex representations with just sparse data?' As stated by Tenenbaum et al. [111], "yet children routinely infer causal links from just a handful of events [45], far too small a sample to compute even a reliable correlation!". As suggested by Plato, another source of information (such as abstract background knowledge) must be available to help this inductive learning process. "Psychologists and linguists speak of constraints; machine learning and artificial intelligence researchers, inductive bias; statisticians, priors." [111]. This abstract knowledge, or over-hypothesis, restricts the hypothesis or the model space at a less abstract level, reducing the complexity to find an appropriate model for explaining the data. Where does this over-hypothesis come from? Some authors suggest that over-hypothesis can be learned simultaneously with the model (the blessing of abstraction as opposed to the curse of dimensionality) using, for instance, hierarchical Bayesian approaches [62] [111][44]. Hierarchical Bayesian approaches are relevant in the context of this work. One of the problems with these techniques is the quantity of data needed to obtain good results. Data is specific to every building and cannot be easily mutualized. It is sparse, and it may be necessary to have the system initially running for a long period to acquire enough data to get an initial usable causal model. As a result, the use of a knowledge model of the environment as a high-level predefined expert knowledge has to be investigated.

2.7 Quality of an explanation

Keil et al.[61] propose three dimensions to evaluate the quality of an explanation: circularity, relevance, and coherence. Others such as Kim et al. refer to the credibility of an explanation [64].

- A circular explanation is an explanation where the conclusion is used as part of the explanation. For instance, "This diet pill works because it helps people lose weight" (extracted from [61]). Complex circularities might be difficult to detect [99].

- Considering an explanation as a speech act, an explanation is relevant for a given goal if it has the appropriate level of detail while not providing unnecessary or unrelated information. Particular care must be taken against ego-centrism as explanations may be based on what we believe the other person knows; the estimation of the other person's knowledge is often extrapolated from our own level of knowledge [88], which could be misleading, resulting in irrelevant explanations.

- Coherence means that the explanation is composed of a set of elements with some of them positively constraining others toward the effects. The credibility of an explanation is related to the causal structure and is called "the causal diversity effect" [64]. A single cause might create several effects that in turn create several effects, and so on, resulting in a causal tree. The causal diversity effect indicates that the further apart two final effects in the tree are, the more they are considered as a good justification of the initial cause. In 1847, the philosopher of science, William Whewell, called this "consilience" [117].

Another concept related to the evaluation of explanations is the Illusion of Explanatory Depth (IOED). The IOED refers to the fact that people usually think, after having received an explanation about a system, that they understand it more profoundly than they really do [101]. Rose et al. have shown that the more people see or visualize a system component, the more they build mental simulations, and the more they believe they understand it.

This presents why it is very difficult to generate the contextual causal explanations in buildings and how to evaluate their quality.

2.8 Conclusion

EMSs are developing in parallel to homes. The various types of EMS vary in their clarity and capability, but the ones capable of analyzing occupant behavior, simulating the environment and allowing optimization of the different parameters are based on energy models. They are, therefore, very complex and unintelligible to the occupants.

To overcome complexity in EMSs, and to make them more comprehensible, there is a need to generate explanations as a way of empowering the occupants by helping them to understand how the system is working and on which principles it is based to make its decisions. Given that, occupants will have confidence in the system as they will understand the logic behind its judgment.

Explanations, which are useful for occupants, have many different forms and levels. "The most intuitive forms most adapted to the EMSs are the causal explanations."

However, finding the causality and extracting it from the models and sensors is still an ongoing scientific problem and many researchers have tried to tackle it from different points of view, as well as trying to improve the quality of explanations to achieve their objective when communicating with occupants.

The next chapter will discuss some different methods usually used for the generation of causal explanations.

Chapter 3

How to generate explanations

As presented in the previous chapters one significant difficulty with model-based energy management systems lies in their opacity i.e. their inability to make clear their reasoning and to justify their recommendations to inhabitants in a convincing manner. To solve this, there is a need to create a persuasive system where the user can trust and cooperate with the system to achieve his goals. This is particularly crucial when one wants to empower the occupants, because energy models are limited and have only partial knowledge about the environment and they do not know the occupant's intentions. The Glass study [43] showed that without trust in the actions and results produced by energy systems as smart agents, their use and adoption as trusted assistants and partners will be severely limited. In the study, they identify and discuss the different methods that can significantly impact user trust in complex systems. They conclude that the availability of explanation capabilities in complex systems, like EMSs, can address the majority of trust concerns.

It is worth reiterating from [47] what has been presented in the earlier chapters, i.e. the need for explanations within expert systems:

- **DO users of intelligent systems want explanations? Why are explanations needed?** It appears that explanations should be provided in smart systems. The situations in which users want explanations are likely to be highly context specific. These situations include the need to resolve perceived anomalies, or a lack of knowledge of use by the intelligent system. Particular tasks such as report production or debugging may also necessitate the use of an intelligent system explanation. In other words, occupants need explanations to understand how energy models are making their recommendations, making sure that their objectives are aligned, and are learning from the mirror service how occupants can adapt their routine to better one.

- **Do benefits arise from the use of explanations? What kinds of benefits?** Explanations in expert system use have been shown to have positive outcomes with better performance, higher user confidence of the system, and, in some cases, improved learning [47].

- **What types of explanation should be provided?** Human beings tend to interpret events with a cause-effect analysis approach [57] [93]. Therefore, causal models are more apprehendable and more accessible to modification [52]; this means they are more easily understood by users [31] [110].

This reinforces the importance of explanations, and the need to transfer knowledge from energy systems to occupants, so that they can trust those systems and understand how they are functioning. If this is not achieved, occupants might neglect their energy systems.

In expert systems, explanations aim to transfer the knowledge of something to someone [68]. According to this definition, to build correct and understandable explanations there is a need to define the appropriate form of knowledge (qualitative and/or quantitative) to communicate with humans and how to obtain this form of knowledge. This will be further elaborated in the next section. The chapter continues with a discussion of Garp3, a workbench for qualitative reasoning. Then it presents the concept of expert knowledge injection, and will end by presenting Bayesian networks, as they are ubiquitous in expert systems in building causal models and generating explanations, and describing how energy models can be enhanced depending on the level of knowledge injected into them.

3.1 Qualitative knowledge

In order to transfer knowledge to users, there is a need to study the existing forms of knowledge. Scientific knowledge can be roughly divided into qualitative and quantitative knowledge. Qualitative knowledge is usually concerned only with symbols, while quantitative knowledge requires recourse to numbers and equations.

With people, two phenomena have been observed: there is great resistance to the use of quantitative knowledge, and quantitative knowledge can be transformed into qualitative knowledge [66].

For example, a beginning high school student might not know that a ball rolling down a (frictionless) hill will reach the same height on the next hill with the same gradient. After learning about potential and kinetic energy, he discovers that the heights must be equal. Now certain kinds of questions can be answered without resorting to equations. The question, "Will the ball make it over the next hill?", can be answered by a simple comparison of the two heights.

Dryllerakis [12] defined five main streams of research that can be identified in qualitative physics today. The motivation in each field is different and normally is the driving force for the development of each theory. The main streams are represented schematically in figure 3.1

Commonsence Reasoning About the World. Bobrow proposed that [12] "Pat Hayes was the first to realize that humans reason about the

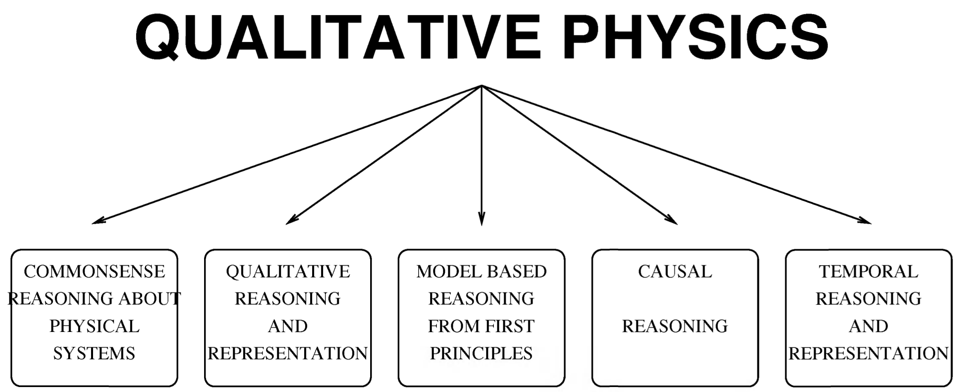


Figure 3.1: The research fields of Qualitative Physics [12]

surrounding world in a natural way: it does not seem to bother them! The part of physics dealing with everyday life phenomena was named naive physics and dealt with commonsense reasoning. Most people, Hayes argued, are able to predict that a ball thrown in the air will eventually drop on somebody's head." This is done without the use of mathematical formulae or the resort to any physical theory. This kind of physical reasoning, performed in the mind of everyone, seems to be much different from the one of a physics student.

Qualitative Reasoning. All scientists have at some stage to reason qualitatively about their domain[12]. This involves qualitative solutions to novel problems as well as guidance in the search for creating a suitable set of equations for solving their problem analytically. Natural and economics sciences often have to deal with vague and imprecise information about variables describing a system. Causal links may also be available between variables but not of a precise form. It is not uncommon that the analytical solution of a set of mathematical equations does not exist. Numerical simulation is often used in this case, a computational approach with much overhead and the need for a correct interpretation of the output.

Model-Based Reasoning From First Principles. Research in this field was initiated by the desire of the academic community to build expert systems without the limitations of the currently available ones. The lack of explicit representation of fundamental knowledge makes the expert systems not so expert, as they cannot predict or advise on any situation not conforming to their if-then rules[12]. A different kind of knowledge needs to be encoded as well as a powerful reasoning mechanism. The great advantage of such an approach, is the possibility of sharing knowledge between overlapping fields. For example [12], if two expert systems exist, one about weather forecasting and one about air-traffic control, it will be almost impossible to share their common knowledge and reasoning processes. In model-based qualitative physics, researchers try to reason from a detailed representation of the domain, the domain model, and the causal dependence of the fundamental quantities. Three main approaches can be seen in this field. The first one is credited to Kleer who in 1984 presented his computer program ENVISION [25]. The program was given a description of the physical system in a pre-specified way and was able to predict behavior using a process called *envisionment*. This term is scattered throughout the literature to express the process of creating a set of possible system transitions over time. In order to overcome the problem of qualitative information, de Kleer presented an algebra of signs which from then on is given the name qualitative algebra. Physical quantities are quantized to a minimal set of designated values that are useful in qualitative reasoning, normally just the signs of the value. The second approach is based on Forbus. He developed a theory of processes called Qualitative Process Theory which he applied to create the program Qualitative Process Engine [41]. In a frame-like environment, the user supplies the basic processes available in the system as well as information about the objects in the system. The third approach is based on qualitative simulation [67]. The behavior of the system is described by a set of qualitative equations and a constraint propagation algorithm tries to solve them over time. It is important to note here that most of the interest in the field was focused on qualitative dynamics. Qualitative dynamics, according to Dryllerakis[12], means representation of the time varying aspects of physical systems in their dynamic behavior. Physical systems are simulated by black boxes described by sets of variables. The dynamic behavior is explored by monitoring the time dependence of the variables. Objects are not represented explicitly but rather through their qualities (the variables).

Causal Reasoning. This is yet another important aspect of qualitative physics. There is little doubt that causality plays a central role in most aspects of human lives[12]. AI programs built in this field are mainly concerned with providing intuitive causal explanations about physical phenomena. These explanations must be based on a knowledge of causal relations in the world. Most expert systems when asked to explain their conclusions will do so by supplying the precompiled heuristics that lead them to the result. A human expert will explain his reasoning based on fundamental causal mechanisms of the domain. Causal reasoning also tries to produce the causal links that will explain the trace of an analytical calculation. In that way a causal explanation of the result can be achieved.

Temporal Reasoning. Temporal reasoning can be understood as the search for a suitable representation of time dependent quantities (they can be database records) as well as a reasoning process for processes involving time[12]. Much work has been done in this field due to its closeness to other more developed fields. Central work for qualitative physics is that by Allen in his interval representation of time [2]. This part of the research is incorporated in qualitative physics as reasoning about physical systems that involves reasoning about time. Allen's approach was adopted as a suitable time representation by researchers in the field as a certain quantization of time is achieved. The time line is split into intervals and reasoning is concerned with time intervals instead of time points.

This section presented a general view of the different fields working in qualitative knowledge and its use. The next section describes Garp3, an information tool built by researchers and engineers to allow qualitative simulations and causal reasoning.

Garp3

Garp3 is an easy to use workbench for qualitative reasoning. It is a workbench that allows modellers to build, simulate, and inspect, qualitative models of systems behavior. The workbench employs diagrammatic representations that enable users to interact with model content and simulation results and provides seamless interoperability between the different contexts [14]. Garp3 is used to create conceptual models in situations where numerical information is sparse or unavailable, or when it is necessary to formalize the theoretical understanding of how systems behave[15]. Garp3 can be applied to stakeholder management or dissemination activities to illustrate and explain phenomena. The workbench can also be used in formal education to have learners express concepts, or interact with existing models, and support them in developing their understanding of 'how things work'.

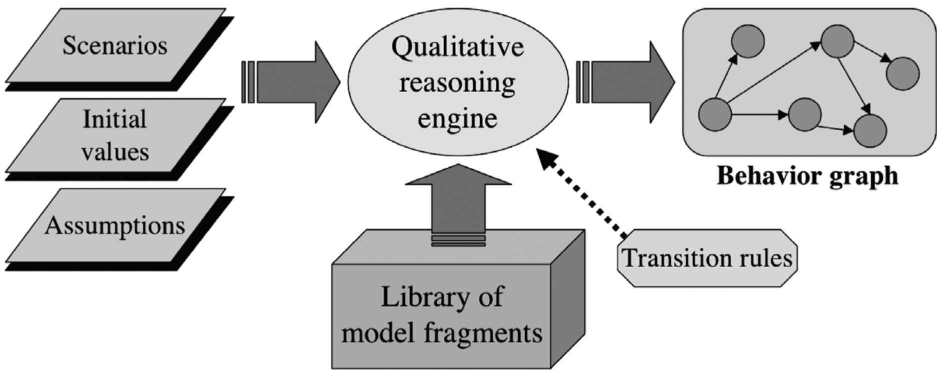


Figure 3.2: Basic architecture of the qualitative reasoning engine [15]

Garp3 incorporates a range of techniques from AI known as knowledge-based techniques. These techniques are going to generate all the possible qualitative states and evolution for the system. Garp3 also applies different aspects of qualitative reasoning like the model-based, causal, and temporal reasoning from the previous section.

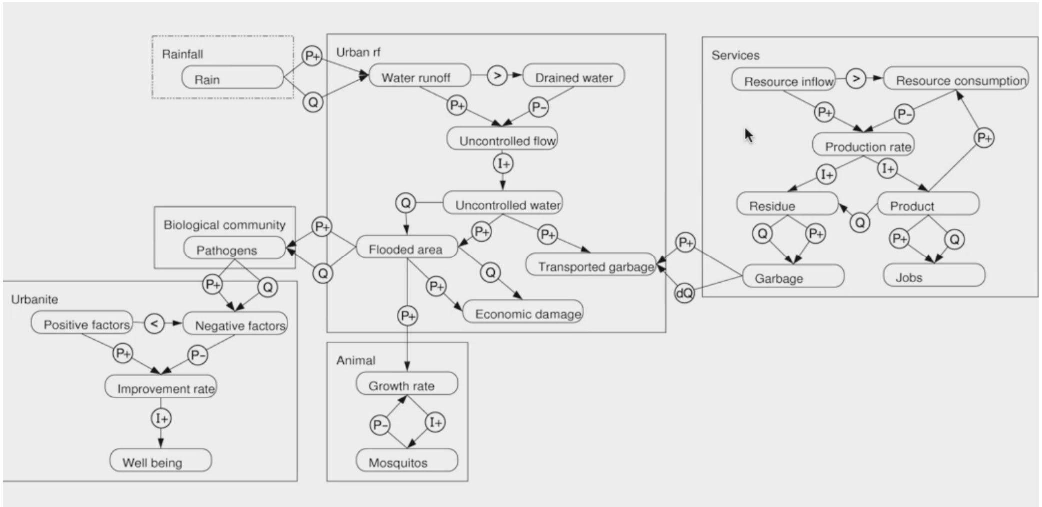


Figure 3.3: Example for modelling Mosquito growth rate with habitat in Garp3 [14]

Figure 3.3 presents an example of Garp3. This example illustrates the need to model each actor and each variable in the environment. It also requires presentation of the different known relations between the variables. For this reason, applying Garp3 to an apartment is very complex as all the appliances, actions, and context variables need to be modelled, and the definition of the different relations between them. Even with a highly simplified model the qualitative simulation took a very long time. The output gives dozen of qualitative states to analyze, like those shown in figure 3.4

To simplify the modelling part and to better organize it, Garp3 presented the "model fragment" concept to enable modellers to build different sub-models. This concept will be clarified further in the next section.

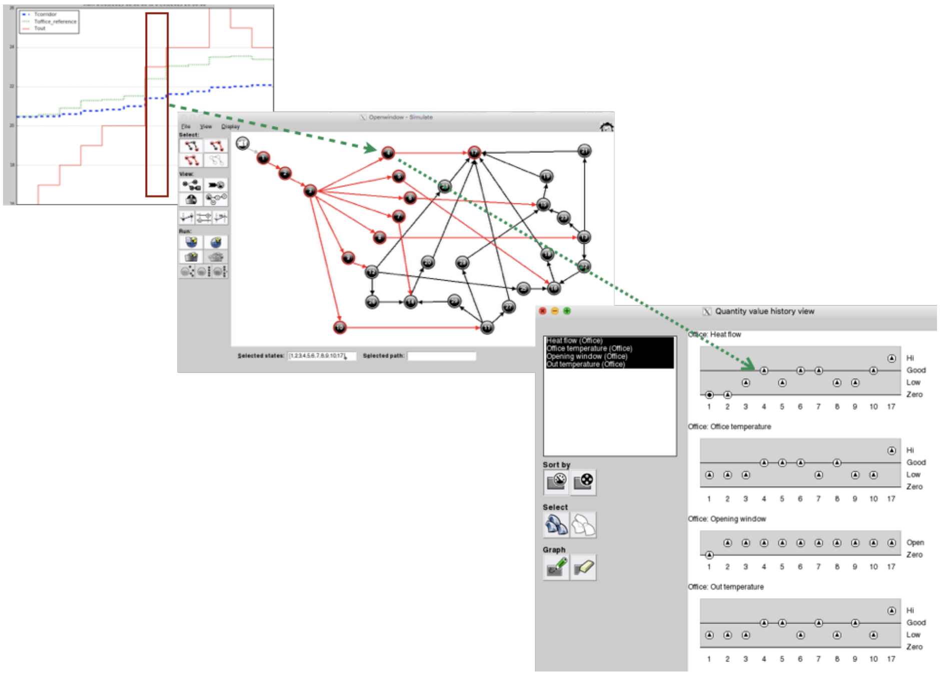


Figure 3.4: Garp3

Model Fragments

Model fragments describe part of the structure and behavior of a system in a general way. They are partial models that are composed of multiple elements. In general, a model fragment has the form of a rule. This means that model elements are incorporated as either conditions or consequences. For example a model fragment can be the relation between the heater, the inside temperature, and the electrical consumption. Model fragments themselves can be reused within other model fragments as conditions, called imported model fragments. Furthermore, sub-classes of model fragments can be created, which augment the parent model fragment with new elements [14].

As mentioned before, Garp3 is a tool that will generate all possible qualitative possibilities for a complex system like an apartment. Even the simplest scenario (a model with a specific context) might have a dozen possible qualitative states.

To implement a Garp3 system to build a general causal model is very difficult, as it demands the valuable time of an expert to define all the limitations and possibilities in the system. On the other hand, Garp3 uses an exciting concept for model fragments. The model fragment concept is going to be very useful in this study. But here, a model fragment is more general as a fragment can be: a sub-model allowing the exploration of different variables, where those variables are not accessible by the model or any other method. It might be a potential causality or a forbidden one that can be determined by an expert or any knowledge injected by an expert like defining the groups of different variables. This will be further elaborated in the next section.

3.2 Knowledge injection

As discussed earlier (in the second chapter), the amount of information that can be transferred to occupants is related to the quantity of knowledge injected into the system. This knowledge comes from the energy models or that injected in the form of model fragments. From 3.5 it can be seen that in the lower level of knowledge (top of the figure) the only knowledge that exists is quantitative coming from measured sets of data.

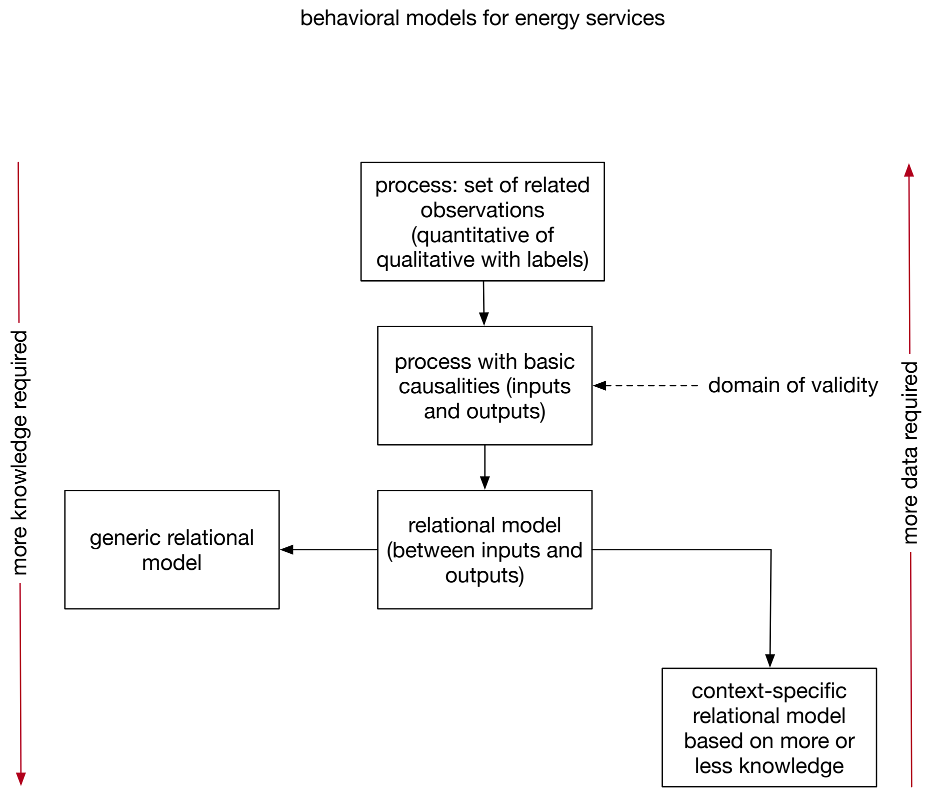


Figure 3.5: Behavioral model for energy services

The next level is when some knowledge is injected. It takes the form of sets of variables with input, intermediate and output variables. The physical variables involved in building simulation models can be grouped into the following categories (figure 3.6), based on the causal relation (depicted by the arrows in figure 3.6) among the physical phenomena they demonstrate:

1. *Occupant actions* ($\mathcal{A}$): This is the set of variables which are directly controllable by the occupants and hence, the system for building energy management can recommend an associated change. For example, in the primary case study, the set $\mathcal{A}$ comprises of opening/closing of doors (ζ_D) and windows (ζ_W), and turning on/off a localized heater (ζ_H).

2. *Physical context* ($\mathcal{C}$): This is the set of variables which cannot be controlled by the occupants. For example, in the primary case study, the set $\mathcal{C}$ consists of factors like outdoor temperature (T_{out}), humidity, illuminance, wind speed, number of occupants ($n(t)$) at t^{th} time quantum, temperature of neighboring zones (T_n), electric power consumption from work-associated routine appliance usage (P_{elec}), and so on.

3. *Intermediate variables* ($\mathcal{I}$): This is the set of variables affected by primary cause variables which represent the physical variables through which some occupant actions, in a certain context, leads to a particular level of occupant satisfaction. For example, in the primary case study, the set $\mathcal{I}$ contains some parameters (I_1), which are estimated through a trained building model, like air flow (Q) and heat flow (φ), along with some parameters (I_2), which are measured through multiple sensors, like indoor temperature (T_{in}) and indoor CO_2 concentration (C_{in}).

4. *Occupant satisfaction assessment* ($\mathcal{S}$): This is the set of variables which represents the effects desired by the occupants for a comfortable lifestyle. For example, in the primary case study, the set $\mathcal{S}$ is represented by indicators of thermal comfort which is for the moment defined between ($21°$ to $23°$)(σ_{temp}), carbon-dioxide concentration (CO_2) based air quality comfort (σ_{air}) which should be as minimal as possible, financial

profit gained by energy savings through optimized usage of heaters (σ_{cost}) and number of changes in recommendations (δ_{WD}).

In general, the occupant actions ($\mathcal{A}$) and the context variables ($\mathcal{C}$) are the potential cause for the changes in the intermediate variables ($\mathcal{I}$) and the occupant satisfaction ($\mathcal{S}$).

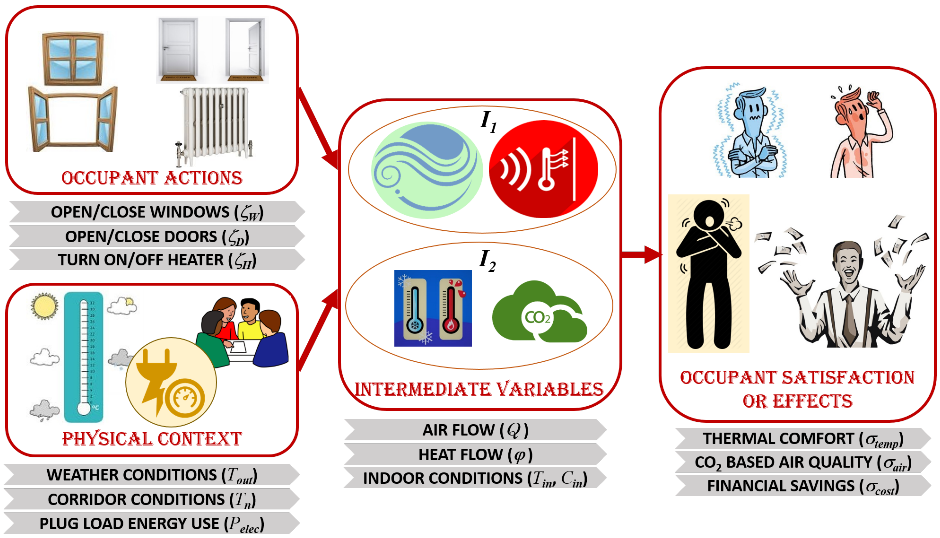


Figure 3.6: General schema with basic causalities

The third level of models (a relational model between inputs and output) model has a representation of the relations between the input variables and the output. This means that by knowing the inputs it is possible to deduce the impact on the output variables. The most precise type of models (context specific relational models) are those in which a full schema can be built between the variables that enable the prediction and the evolution of each variable within a different context and different input.

3.3 Causal graphs

Montmain [81] described these by: "A causal graph describes explicitly the unidirectional relationships between variables. It is a directed graph, often abbreviated as a digraph: the nodes symbolize the variables, and the arcs the relations between them." Modelling causality through graphs brings an appropriate language to describe the dynamics of causality. Whenever an event A is a cause of B, an arrow is drawn in that direction ($A \rightarrow B$).

Causal graphs are known and used in different domains. For example, Gentil[42] presented how they use causal graphs in the domain of diagnostics to help and cooperate between supervisors and expert systems. Causal graphs are also used in the medical and economics domains between authors. Bayesian networks are an example of causal modelling in a situation where understanding is important; they will be discussed further in the next section.

3.4 Bayesian networks

The Bayesian network (BN) consists of an acyclic directed graph (ADG) whose nodes represent random variables, together with a conditional probability distribution for each node Xi given its parents,$P(xi|pa(xi))$. The conditional probability for a node without parents is just its prior probability $P(xi|\Phi) = P(Xi)$. These probabilities can be obtained from statistical data (for instance, from measured data), from the literature on the specific domain or by the judgment of human experts. The joint probability represented by a Bayesian network is:

$$P(x_1, ..., x_n) = \Pi_i P(x_i|pa(x_i)) \tag{3.1}$$

A Bayesian network is an annotated directed graph that encodes probabilistic relationships among distinctions of interest in an uncertain-reasoning problem [92]. The representation formally encodes the joint probability distribution for its domain, yet includes a human-oriented qualitative structure that facilitates communication between a user and a system incorporating the probabilistic model.

This distribution satisfies the d-separation property [92] and its equivalent, the Markov property [55][87], which states that a node is independent of its non-descendants in the graph given its parents. Roughly speaking this property implies that a link X Y in a Bayesian network represents a probabilistic dependence between X and Y, while the lack of a link represents probabilistic independence. A finding is a piece of information that states with certainty the value of a random variable[68]; a finding may be, for example, that the patient is a male; other findings might be that he is 54 years old, that he has a fever, that he does not usually have headaches, and so on. The set of findings is called evidence, e. Probabilistic reasoning consists of computing the posterior probability of the unobserved variables given the evidence; for instance, $P(xi|e)$ or $P(xi, xj, Xk|e)$. This process is usually called evidence propagation and is based, more or less explicitly, in the application of Bayes theorem.

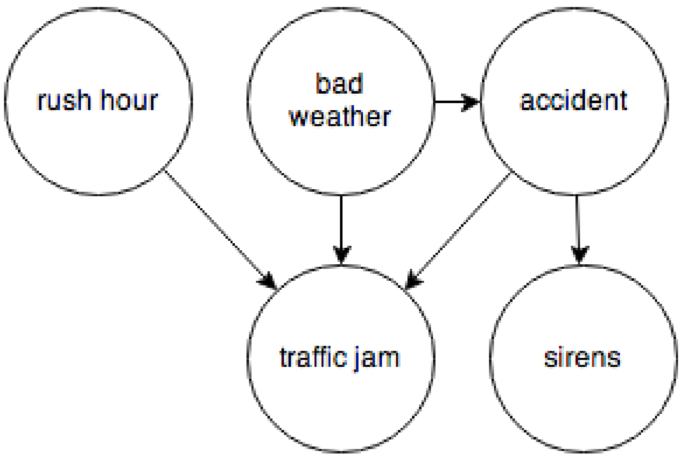


Figure 3.7: Example of a Bayesian network

From a mathematical point of view, a BN is just a model for representing probabilistic dependencies and independencies; in this case, a link, considered by itself, has no meaning. However, when a BN is built as a model of a real world system, a link A B is causal when A is a cause of B, i.e., when there is a mechanism by which the value taken on by A influences the value of B . A BN is said to be causal when all of its links are causal. Causality and probability are closely related, because causality normally implies a pattern of probabilistic inter-dependencies, which provides clues about causality[75]. In fact, a necessary condition for establishing the presence of causality is statistical correlation [33].

Causal Bayesian networks support certain qualitative reasoning patterns, which can be identified in order to explain the results of inference [32].

In a similar case to Garp3, building a general causal model for an apartment using Bayesian networks is very complex. This is due to the large number of variables in each apartment and the changing causalities with different contexts.

3.5 Conclusion

This chapter studied the quantitative and qualitative forms of knowledge and their different representations. It presumed that qualitative knowledge is the best form in which to transfer a piece of knowledge to inhabitants. Qualitative knowledge is usually enough to give the user the ability for good reasoning about energy systems. The measured, or simulated, data can be used to confirm the reasoning about the energy system.

Then, the chapter presented Garp3 which is an information tool to allow researchers to do qualitative simulations for different systems. The reasons for the difficulties of using Garp3 to build a general causal model for buildings have been described. The concept of "model fragment" proposed by Garp3 was explained in this study and the chapter presented how model fragments can be used to inject some information into the EMSs. This injected knowledge is important and can be useful to occupants.

Finally, the chapter presented causal graphs and Bayesian networks. It discussed some of their important properties, which will be helpful in the following chapters.

Chapter 4

Explanations generation with knowledge models

Table 4.1: Description of the symbols used

Parameters	Meaning	Remarks
τ	Average temperature of the building envelope	
R_n, R_{out}, R_W, R_D	Thermal resistance of neighboring zones, outdoor, window and door	
R_i, C_i	Equivalent resistance and capacitance representing inertia	Data from $\mathcal{H}$
$R(\zeta_D, \zeta_w)$	Equivalent resistance	By equation (4.3)
ζ	Represent the different possible actions	
ζ_D, ζ_w	Represent door and window opening	
T_{in}, T_n, T_{out}	Temperatures inside, with adjacent corridor and outside	
φ_{in}	Total indoor energy gains	
ρ_{air}	Air density	Typical value is 1.204m^3
$c_{p,air}$	Specific heat of air at room temperature	Typical value is 1.004 kJ.kg^{-1}.K^{-1}
C_{in}, C_n, C_{out}	CO_2 concentration indoor, with neighboring zone and outdoor	$C_{out} = 395 \times 10^{-6}$ mol per mol of air (constant)
Q	Represent the different air and heat flow	
Q_n, Q_{out}, Q_W, Q_D, Q_{WD}	Air flow with adjacent corridor, outdoor, through window, through door, through window and door (cross-ventilation)	
S_{CO_2}	Breath production in CO_2 from each occupant	Typical value is 8.73×10^{-6} mol.m^{3}s^{-1} per person per mol of air
P_{elec} or $\varphi_{appliances}$	Power drawn from electric supply or net heat flow from appliances	
P^{max}_{heater}	Maximum energy consumption associated with water circulation for hourly heater usage	Typical value is 2000W

At the most fundamental level, smart buildings deliver useful building services. They aim to give occupants a good level of comfort (for example: thermal comfort, air quality and more) at low cost and low environmental impact. To make this possible requires not only developing intelligent energy systems like the ones presented in the second chapter but also needs the involvement of occupants in the loop together with the EMS and the building environment. Occupant activities play a significant role in building energy consumption. Yet, EMSs are not designed to cooperate with occupants. The embedded knowledgein the EMSs cannot be easily shared with occupants because they are designed and made by engineers with the objective of optimizing the energy consumption. A gap has been created between EMSs and occupants because of this lack of cooperation. The occupant knows his intentions and goals, but he does not know how to realize them. While the EMS knows how to optimize different objectives, it is not always aware of the occupant's intentions. For example, if the system recommends opening the window to enhance the air quality the inhabitants may not want to do that because there is a lot of noise outside, and the system is not aware of that. Similarly the inhabitant does not know the impact of not opening the window on air quality or if there are any different solutions to satisfy his comfort without opening the window. Consequently, it is important for occupants to interact and cooperate with the energy system. To empower occupants, there is a need to build a bridge of cooperation between them and their EMS. This bridge should allow occupants to understand which criteria the EMS is acting upon when making decisions. Additionally, it should provide occupants with feedback about the impact of their actions on energy consumption and their comfort criteria. This bridge of cooperation can be built using explanations, as a powerful and intuitive tool to transfer knowledge with the expert systems as mentioned in the second chapter. This chapter proposes two approaches to generate causal explanations from EMSs. These two approaches use a model-based EMS (knowledge model). The next chapter focuses on how to generate the explanations without a knowledge model.

The first form of explanations compares two scenarios (consisting of actions, context, and effect), like for example, what the user did in a day, and a recommended scenario generated by the EMS. Then it simulates the different possible actions using the energy models to estimate the impact of the user actions and generate "differential explanations".

The second form is based on Bayesian networks and uses the different possible simulations to generate the "direct explanations".

<table>
<tr><th></th><th>With knowledge models</th><th>Without knowledge models</th></tr>
<tr><td>Differential explanations</td><td rowspan="2">This chapter</td><td rowspan="2">Next chapter</td></tr>
<tr><td>Direct explanation</td></tr>
</table>

The chapter starts by stating the problem and by describing a case study from which data are collected, and where the approach is deployed for testing. Then, it presents the energy model and its different variables. Finally, it describes the steps to obtain the explanations.

4.1 Problem statement and general solving principles

This chapter tackles the problem of generating causal explanations with a model-based EMS, as an example of knowledge models, to involve occupants in the loop with their EMS. Due to the complexity and mathematical formalism of the knowledge models, they are not suitable for interactions with inhabitants: the intrinsic knowledge they contain is not directly intelligible. They were built by engineers to predict and simulate the environment, in order to optimize the energy consumption and not to interact with occupants. Differential explanations are based on comparing two different scenarios. But when is it possible to compare these scenarios? For instance, two scenarios with different contexts can be compared technically, but the results might not have much meaning for the end user because the approach cannot separate the effect from different actions when the context changes. For example opening the window will affect the inside temperature depending on the outside temperature and other variables. For that, the scenarios should have the same context or a comparable one, to be able to get the exact impact of each action or the impact of the context changing.

Additionally, there is a need to define the cause-effect relationships between the variables. It is the objective to generate causal explanations.

Buildings have inertia which implies that actions have delayed effects. How can we identify the actions that have those delayed effects?

The next section presents a case study, and the chapter continues with the different steps to generate explanations using the testbed as an example.

4.2 Description of experimental testbed

For this work, an office at Grenoble Institute of Technology was studied, where four researchers work. The descriptive features of the office are shown in Figure 6.11 and are listed as follows:

- *Dimensions:* The office room has an approximate floor surface area of 31.26m^2 and a height of 2.50m. Thus, the volume of the room (V) is 78.15m^3. The outer wall has a thickness of 0.30m and the inner plaster wall has a thickness of 0.06m.

- *Aesthetics:* The room walls are white painted and the windows are south-east facing for proper illuminance and availability of ample natural light. There are two potted plants placed at diagonally opposite corners of

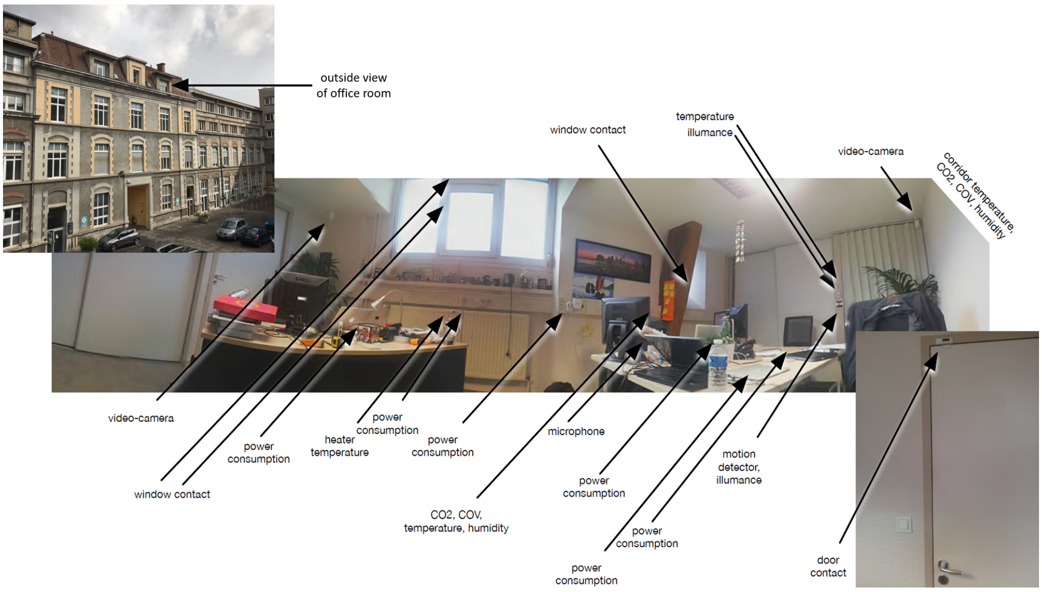


Figure 4.1: Office at Grenoble Institute of Technology fitted with 27 sensors

the room. There are two tables and five chairs (for the four researchers and one guest) which is the only furniture present in the room, apart from the mobile heater.

- *Location:* The road next to the building wing where the office is located, is a narrow, quiet street. The window-facing side of the building is open and has a parking place in front of it. Thus, nothing blocks the natural air and light flow. Considering all these aspects, the effective solar factor from different windows turns out to be around 17.29%.

- *Attached sensors:* The office and its adjacent corridor are fitted with a total of 27 sensors for recording window opening, door opening, illuminance, acoustic pressure, indoor and corridor physical variables like temperature, humidity, CO_2 concentration and volatile organic compounds (VOC). For office occupancy estimation, the approach of [6] is used, which extends over [7] and validates the results using a video camera. This array of sensors assists in remotely sensing a substantial portion of the physical context of the room.

- *Electricity supply*: Although buildings in France draw power from both variable tariff power supply sources and fixed tariff power supply sources, the selected office draws power only from fixed tariff power supply sources at the rate (E_{elec}) of 0.15 Euros per kilowatt-hour (kWh). The electric power supply helps in running work-related appliances such as laptops and projectors. Thus, the approach used in this work can be extended to buildings in other countries where variable tariff power supply schemes have not yet been introduced.

- *Heater*: The heater in the room exchanges heat between the room air and the hot water circulating in the pipes. Although the water circulation is managed centrally by such buildings, yet the proportional zonal expense, incurred on the fuel consumed to heat the equivalent amount of water circulated over the entire duration of heater usage. For the selected testbed, this fuel consumption associated expenditure (E_{fuel}) is at the rate of 0.089 Euros per kWh.

Other behavioral adaptations and activities dictating the metabolism of the occupants (φ_{bodies}), and in turn influencing their comfort, are assumed to be nearly the same over the data collection period to ensure identical treatment and to discard other kinds of bias. The typical value of such bodies' metabolism (φ_{bodies}) is 80 watt(W) per person [83].

Sensor data and weather conditions were recorded for a consecutive period of about one and a half years (April 2015 to October 2016) at hourly intervals and stored in the historical database ($\mathcal{H}$) for future simulations (using physical models). Primarily, the historical database ($\mathcal{H}$) has two purposes as follows:

1. To tune the physical model parameters [106] by minimizing the standardized root mean square error between the simulated indoor conditions and measured indoor conditions, under the same context, by replaying the historical actions stored in the database,

2. To provide the contextual variables for simulating indoor conditions corresponding to hypothetical actions.

4.3 Physical knowledge model for building performance simulation

Identifying the proper physical knowledge model is a challenge of the proposed framework. Developing the model of the building is a difficult work as there is a multitude of ongoing physical phenomenon (building-occupant-environment interactions). The identified significant physical interactions involve heat flow and air flow affecting the indoor temperature and indoor CO_2 concentrations. When the desirable indoor quality (inside temperature, air quality) could not be attained, a heater was used to deal with the situation [89] and hence, energy consumption (and its associated expenditure) is also modelled.

4.3.1 Thermal model

Several thermal models for this office have been studied in [106] from which a model with one capacitor was used for this work. The equivalent model is represented in Figure 4.2a and is described by equations (4.1) to (4.3), the parameters of which are described in table 4.1. Besides the physical context, the indoor temperature (T_{in}) heavily relies on occupant actions of opening or closing of doors (ζ_D) and windows (ζ_W).

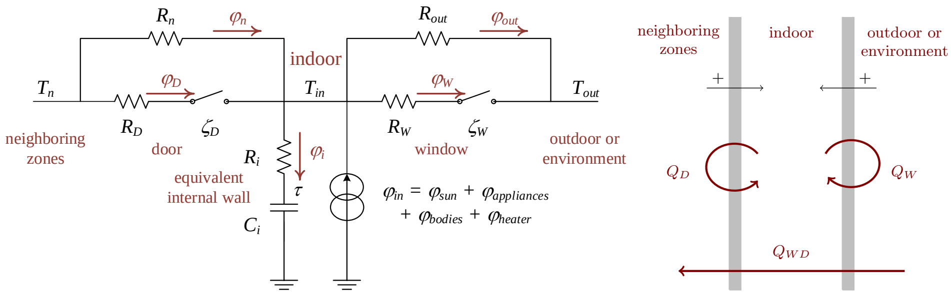


Figure 4.2: Building simulation models: (a) thermal model [106], (b) aeraulic model [74]

$$
\begin{aligned}
\frac{d\tau}{dt} =& \frac{R(\zeta_D, \zeta_W) - R_i}{R_i^2 C_i}\tau + \frac{R(\zeta_D, \zeta_W)}{R_i C_i}\left(\frac{1}{R_{out}} + \frac{1}{R_w(\zeta_w)}\right) T_{out} \\
&+ \frac{R(\zeta_D, \zeta_W)}{R_i C_i}\varphi_{in} + \frac{R(\zeta_D, \zeta_W)}{R_i C_i}\left(\frac{1}{R_n} + \frac{1}{R_D(\zeta_D)}\right) T_n
\end{aligned} \tag{4.1}
$$

$$
\begin{aligned}
T_{in} =& \frac{R(\zeta_D, \zeta_W)}{R_i}\tau + R(\zeta_D, \zeta_W)\left(\frac{1}{R_{out}} + \frac{1}{R_W(\zeta_w)}\right) T_{out} \\
&+ R(\zeta_D, \zeta_W)\left(\frac{1}{R_n} + \frac{1}{R_D(\zeta_D)}\right) T_n + R(\zeta_D, \zeta_W)\varphi_{in}
\end{aligned} \tag{4.2}
$$

$$
\begin{gathered}
\text{where, } \frac{1}{R(\zeta_D, \zeta_W)} = \frac{1}{R_i} + \frac{1}{R_{out}} + \frac{\zeta_W}{R_W} + \frac{1}{R_n} + \frac{\zeta_D}{R_D}, \\
R_D(\zeta_D) = \frac{1}{\rho_{air} c_{p,air} Q_D}, \; R_W(\zeta_w) = \frac{1}{\rho_{air} c_{p,air} Q_W}, \\
Q_W = Q_W^0 + \zeta_W Q_W^1 \text{ and } Q_D = Q_D^0 + \zeta_D Q_D^1 \\
\text{with time-invariant } R_n,\ R_{out},\ R_i \text{ and } C_i.
\end{gathered} \tag{4.3}
$$

4.3.2 CO_2 based air quality model

Several factors influence the indoor air quality, like humidity, odor, crowding, and CO_2 concentration. As prolonged exposure to high concentration of CO_2 is not recommended [104], this work concentrates on the CO_2 based air quality and uses the aeraulic model recognized in [105]. The equivalent model of the office is represented in Figure 4.2b and is described by equations (4.4) to (4.6), the parameters of which are described in table 4.1. Given the physical context, the occupancy ($n(t)$) estimation approach of [6] is considered. Thus, when the other parameters have been learned, this aeraulic model can simulate the indoor CO_2 concentration based on occupant actions of opening or closing of doors (ζ_D) and windows (ζ_W).

$$\begin{aligned} V\frac{dC_{in}}{dt} =& -\left(Q^0_{out} + Q^0_n + \zeta_W(t)Q_W + \zeta_D(t)Q_D\right) C_{in}(t) \\ &+ \left(Q^0_{out} + \zeta_W(t)Q_W\right) C_{out} \\ &+ \left(Q^0_n + \zeta_D(t)Q_D\right) C_n(t) + S_{CO_2}n(t) \end{aligned} \tag{4.4}$$

$$\begin{aligned} C_{in}(t) =& C_{out} + \frac{Q^0_n + \zeta_D(t)Q_D}{Q^0_{out} + Q^0_n + \zeta_W(t)Q_W + \zeta_D(t)Q_D} C_n(t) \\ & \frac{S_{CO_2}}{Q^0_{out} + Q^0_n + \zeta_W(t)Q_W + \zeta_D(t)Q_D} n(t) \end{aligned} \tag{4.5}$$

$$\begin{aligned} \text{where, } Q_{out}(t) &= Q^0_{out} + \zeta_W(t)Q_W \\ \text{and } Q_n(t) &= Q^0_n + \zeta_D(t)Q_D \end{aligned} \tag{4.6}$$

4.3.3 Energy cost model

Since the energy cost model for the office had not been extensively studied before, a preliminary model was used in this work to estimate the energy consumed by the office appliances. There are two sources of energy consumption: (i) the power drawn from the electric supply (P_{elec}) which is described in table 4.1, and (ii) the energy generated by fuel consumption to heat water for circulating in the heater pipes (P_{fuel} as shown in equation (4.7)). Considering both the factors (P_{elec} and P_{fuel}), the overall energy consumption related expense is indicated by σ_{cost}.

$$P_{fuel}\left(t\right) = \zeta_H\left(t\right) \times P_{heater}^{max} \tag{4.7}$$

where ζ_H represent the opening percentage for the heater valve.

It should be noted that in the testbed, being an office environment, it is difficult for the occupants to precisely control the electric power drawn by the work-related appliances in order to minimize the associated expense. However, the heater (ζ_H) can be controlled by the occupants, not only for minimizing the associated expenditure but also for balancing indoor physical variables (T_{in} and C_{in}). Thus, ζ_H has an impact on ζ_W and ζ_D as well, and the important query is now to explore the ζ_W, ζ_D and ζ_H for optimal effects.

4.3.4 Encoding the solution vector

The occupants can perform actions that can affect the building system. They might be like opening or closing of doors (ζ_D) and windows (ζ_W) and turning on or off a heater (ζ_H) (in the winter). This work considers an hourly granularity (sampling time is one hour) so the recommended values for each of these actions were explored at an hourly interval. As there are three variable actions, each of which has 24 samples corresponding to 24 hours of a day, the actions vector is 72-dimensional where $72 = 24$ samples $\times 3$ actions as shown in Figure 4.3.

It should be noted that vectors ζ_W, ζ_D and ζ_H can be real-valued (non-binary) as per the models described in section 4.3. Indeed the values of these variables for the actions in $\mathcal{H}$ are real-valued. Yet the interpretation of these variables is ambiguous if the recommendations are in terms of real-values. For example, $\zeta_W^k = 0.25$ does not mean partially opening the window by 25%

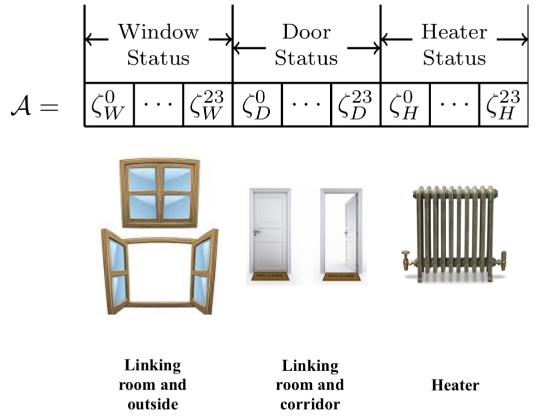


Figure 4.3: Representing a 72-dimensional actions vector

for the k^{th} hour. It means fully opening the window for a span of 15 minutes (25% of an hour) during k^{th} hour.

The next section will describe the generation of explanations using the described energy models and data.

4.4 Generating explanations

After describing the model-based EMS, figures from 4.4 to 4.7 present different outputs of the system for 5 May 2015 throughout the day from 8 a.m. to 8 p.m. (normal working hours for the office). In Figure 4.4, it can be seen on the left the different window actions registered on that day and the recommended actions generated by the system for that day (window-opening-best); on the right there are the same actions but for the door opening. Figure 4.5 represents the different simulated intermediate variables (heat flow on the left and air flow on the right). Figure 4.6 presents the solar radiation on the left and the estimated occupancy on the right. Figure 4.7 presents the different inside temperature, outside temperature, corridor temperature, and the best temperature simulated when the occupant follows the recommended actions.

The same on the right for the air quality(represented in CO_2 concentrations).

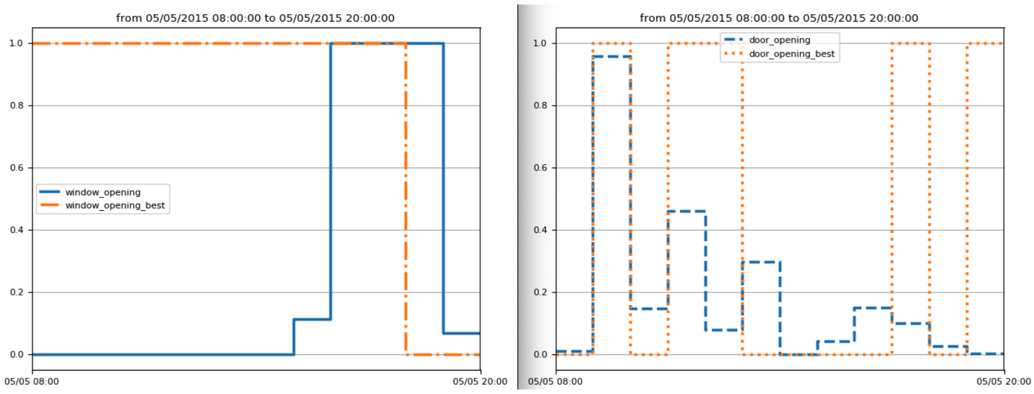


Figure 4.4: window (left) / door(right) opening (measured and recommended)

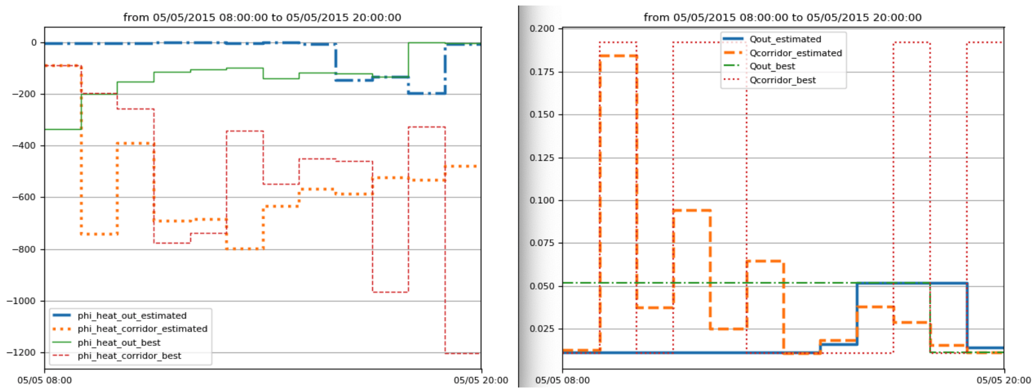


Figure 4.5: Different estimations of the heat flow (left) and air flow (right)

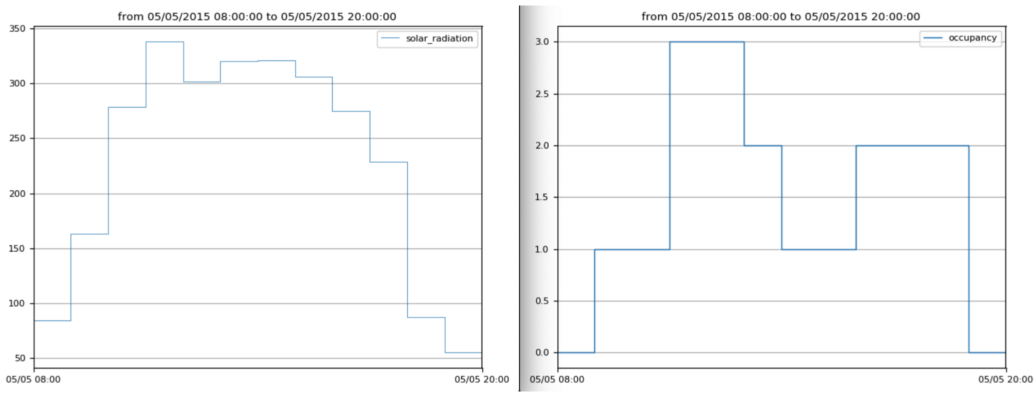


Figure 4.6: Solar radiation (left) and estimated occupancy (right)

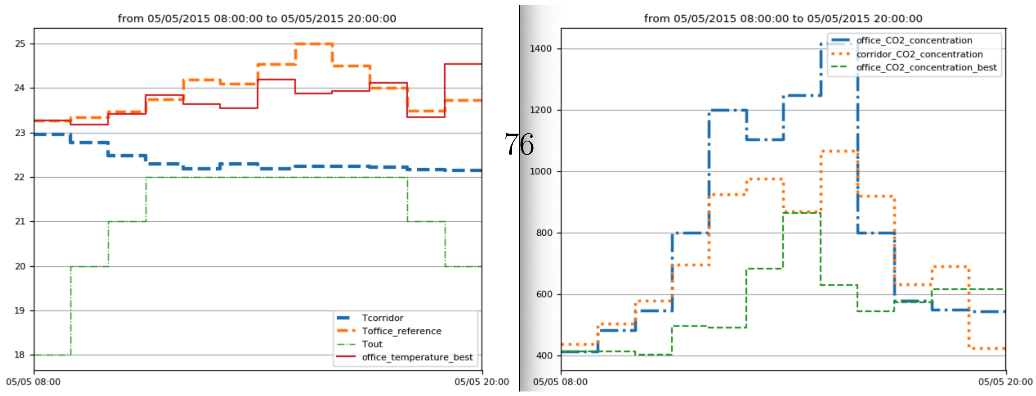


Figure 4.7: Outside, corridor, inside, best (with recommended actions) temperature (left) and corridor, inside, recommended air quality (right)

This example, figures from 4.4 to 4.7, is measured and simulated in a simple environment (one thermal zone): the office. This example presents:

1. The difficulty for the user to understand his environment for one day, because he needs to know and correlate the different variables present in the environment and to understand how they are impacting his comfort criteria (here, the effects).

2. The importance of the occupant's actions. The simulated effects in Figure 4.7 shows how two simple actions like opening the door and window have a considerable impact on the comfort criteria.

3. The cognitive dissonance problem (between occupant's goals and actions). Inhabitants may act in contradiction to their goals (their comfort criteria, cost ...) because they do not understand the impact of their actions. In this example the occupant's goal is to maintain a good level of air quality, yet he closed the window in the morning contrary to what he should have done to reach his objective.

This shows why it is very important for occupants to understand the impact of their actions and here explanations can be very helpful in doing that.

The next section describes differential explanations and how they are generated.

4.4.1 Differential explanations

Differential explanations are constructed by analyzing the difference between two scenarios. Scenarios can be measured, simulated or imagined; they can be set in the past, the present or future. For example, the occupants can compare a past day's actions with an imaginary plan of actions and learn the impact of those actions on their comfort criteria. The scenario can also be a comparison between what the user has done and the recommended actions generated by the EMS to see what he can gain if he applied the recommended actions, as in Figure 4.9. This comparison can also be done between recommended actions for a future day and what the inhabitant likes to do, or simply any plans of actions for a day. This comparison with the recommended actions is important as it could play a role in persuading occupants to change their behavior and follow the recommended one as demonstrated in the first chapter. The comparison includes the set of actions, the intermediate variables, and the effects. The variables are concerned with the consequences of the difference between the user's actions and the recommended actions, for example, and form the basis for the explanations.

Effect variables are impacted by changes in the actions and the context variables. To clearly understand the difference between occupants' different actions, scenarios have to have a similar context to enable comparisons between actions and effects. Otherwise the different effects may be related to the difference in the context variables and not the occupants' actions.

To reduce the risk of false causal relationships or circular explanations, the available variables are classified into four groups: actions, context, intermediate, and effects variables, as shown in Figure 4.8, Natural tendency of causal relations between the groups of variables depending on their role. This is considered as expert abstract knowledge [111].

In Figure 4.9, the differential explanations are illustrated in a table where the first column represents the difference of occupants actions with what the occupant should have done according to the recommended plan, as shown in Equation (4.8).

$$A_k^\star - \tilde{A}_k = \Delta A_k \tag{4.8}$$

where $A_k^\star$ represent the action calculated by the energy model (or any other scenario) at instant k while $\tilde{A}_k$ represents a measured occupant's action at

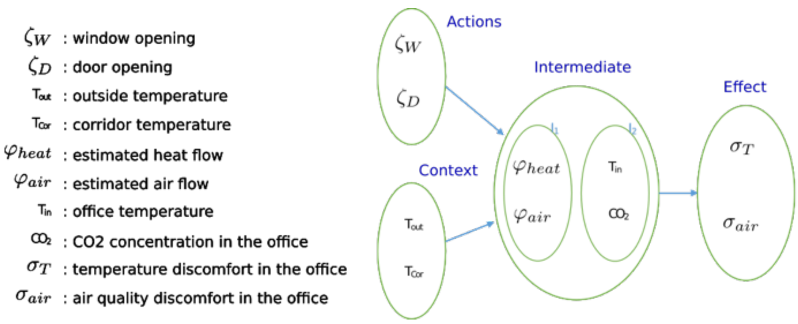


Figure 4.8: General schema to generate explanations

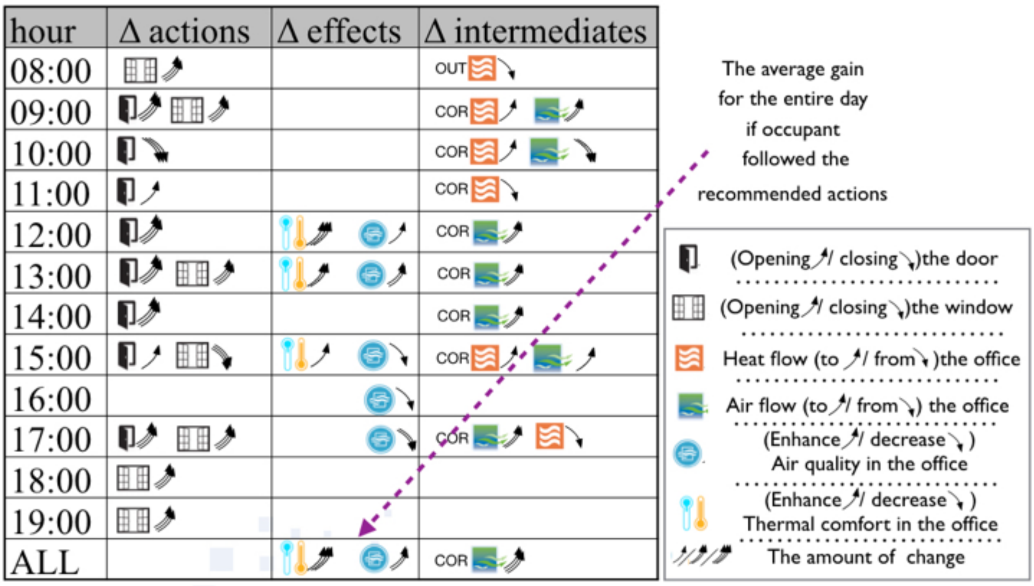


Figure 4.9: Differential explanations through difference between the historical scenario and a Pareto-optimal scenario for 05-May-2015

the same instant. The variable, k, can take any integer value representing hours in the day. In this example, $k = 8, ..., 20$, because it focuses on the time where occupants are potentially present (daytime period). At 8 a.m., for instance, the inhabitant should have opened the window more. At 4 p.m., the user behaves according to best scenario.

The second column presents the effects (Figure 4.8), like in the thermal comfort and the air quality. This is given by Equation (4.9).

$$E_k^{\star} - \tilde{E}_k = \Delta E_k \tag{4.9}$$

where $E_k^{\star}$ represents the calculated effect by the system at instant k while $\tilde{E}_k$ represents the measured effect. The right-hand side of Equation (4.8) and (4.9) denotes the difference in actions of the occupants and the resulting difference in the effect at the k-th instant, respectively.

To better represent the causality in the explanations and extract the knowledge from the system, the intermediate variables are added in the third column of Figure 4.9. Those variables are extracted to render the knowledge from the system explicitly.

The last row, labelled *ALL*, represents the overall gain or loss in the comfort criteria throughout the day, to generate a small summary and give the inhabitant an indicator of their enhancement in general, for the entire day if they follow the recommendations.

When computing the differential explanation, it is necessary to transform quantitative variable values into qualitative ones for a better understanding by the occupants and to define the qualitative distance. For instance, telling the occupant that closing the door at 2 p.m. will cause a large decrease in the airflow and that he will obtain a significant decrease in the air quality level is easier to understand than telling him that a difference in airflow of 30% will lead to a difference in CO_2 concentration of 400 ppm. The transformation from quantitative to qualitative data here is done by dividing the value domain of a variable into 7 sub-domains (3 positive, 3 negative and 1 no-change levels). Those levels were chosen from human feelings according to their impact on the occupants.

The levels for variations in thermal dissatisfaction are given by:

$$\Pi^{T}_{-0.25,-0.15,-0.05,0.05,0.15,0.25}(\Delta\sigma^{k}_{T}(T_{in}))$$

The levels for the variations in air quality dissatisfaction are given by:

$$\Pi^{C_{CO_2}}_{-0.2,-0.1,-0.05,0.05,0.1,0.2}(\Delta\sigma^{k}_{air}(C_{in}))$$

The levels for the variations in the opening of the door and the window are given by:

$$\Pi^{opening}_{-0.7,-0.5,-0.2,0.2,0.5,0.7}(\Delta\zeta_{D})$$

$$\Pi^{opening}_{-0.7,-0.5,-0.2,0.2,0.5,0.7}(\Delta\zeta_{w})$$

The arguments of each of these discretization functions describe the difference of the measured quantity with the proposed optimal value of the quantity.

Except for the no-change level, where arrows are omitted, 1 to 3 arrows have been used to represent the associated sign of variation (arrows direction) and intensity (number of arrows). For instance, in Figure 4.9, the logo of window with three adjacent upward arrows means that the occupant should have opened the window for a much longer period of time during the corresponding time period. Algorithm 1 presents the different steps needed to obtain the differential explanations.

It can be seen that the differential explanations are much easier to understand than the analysis of the 13 plotted curves (Figures 4.4 to 4.7) where the inhabitant has to correlate the different actions, effects and intermediate variables. With a differential explanation, it is easy for an occupant to identify the actions that need to be modified, and monitor the difference gained with respect to different criteria while at the same time using the intermediate variables as elements of understanding.

Algorithm 1 Tabulating differential explanations

Input:

Scenario 1: $\tilde{\mathcal{A}}$, $\mathcal{C}$.
Scenario 2: $\mathcal{A}^\star$, $\mathcal{C}$.

Output: $\mathcal{T}$: table for differential explanations

Use physical model to get $\tilde{\mathcal{I}}$ and $\tilde{\mathcal{S}}$: $\tilde{\mathcal{A}}, \mathcal{C} \xrightarrow{\tilde{\mathcal{I}}} \tilde{\mathcal{S}}$
Use physical model to get $\mathcal{I}^\star$ and $\mathcal{S}^\star$: $\mathcal{A}^\star, \mathcal{C} \xrightarrow{\mathcal{I}^\star} \mathcal{S}^\star$
for $k = t_{start}$ to t_{end} **do**
    $row = k - t_{start} + 1$
    $\mathcal{T}_{row,1} \leftarrow k$
    Obtain $\zeta^{\star,k}$ from $\mathcal{A}^\star$ Different actions at instance k
    Obtain $\tilde{\zeta}^k$ from $\tilde{\mathcal{A}}$
    Calculate $\Delta\zeta^k = \zeta^{\star,k}$ - $\tilde{\zeta}^k$
    Obtain $\sigma^{\star,k}$ from $\mathcal{S}^\star$ obtain user satisfaction
    Obtain $\tilde{\sigma}^k$ from $\tilde{\mathcal{S}}$
    $\Delta\sigma^k = \sigma^{\star,k}$ - $\tilde{\sigma}^k$
    Obtain $Q^{\star,k}$ from $\mathcal{I}^\star$
    Obtain $\tilde{Q}^k$ from $\tilde{\mathcal{I}}$
    $\Delta Q^k = Q^{\star,k}$ - $\tilde{Q}^k$
    $\mathcal{T}_{row,2} \leftarrow$ Qualitative transformation of $\Delta\zeta^k$
    $\mathcal{T}_{row,3} \leftarrow$ Qualitative transformation of $\Delta\sigma^k$
    $\mathcal{T}_{row,4} \leftarrow$ Qualitative transformation of ΔQ^k
end for
$\mathcal{T}_{(row+1),3} \leftarrow$ Qualitative transformation of average values of $\Delta\sigma$
$\mathcal{T}_{(row+1),4} \leftarrow$ Qualitative transformation of average values of ΔQ

4.4.2 Differential explanations with contextual causality

The differential explanation is yielding a list of behavior modifications (opens the door for a longer period, for instance) with the associated impact. However, there are two limitations with such descriptions.

First, there is not a direct link between an action modification and its impact. Buildings have inertia i.e. energy dynamically stored in their structure. This inertia causes a delay and has a smoothing effect on different changes in the building preventing a rapid degradation or augmentation in temperature. Inertia is also present in the room volume for the CO_2 concentration. Thus, occupant actions might have a delayed impact.

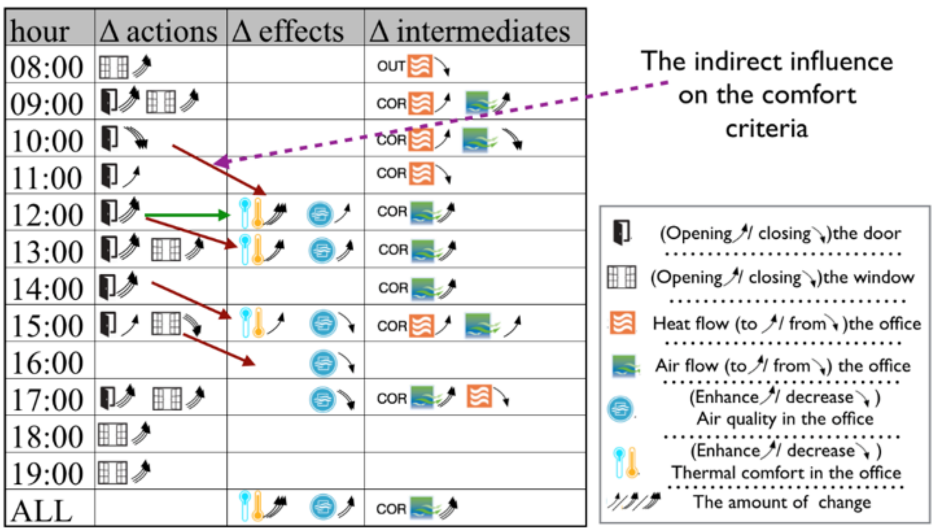


Figure 4.10: Differential Explanations with contextual causality

In Figure 4.10, closing the door at 10am does not have an immediate impact, but it does have a strong impact on the air quality at 12 p.m.: it is a calculated delayed impact.

Second, not all the proposed action modifications have the same importance; some of them have a limited impact and could be skipped if necessary (the inhabitant might not want, for instance, to interrupt his current activity to close the window). But some of them should be followed because of their high impact on the selected criteria (like the previous door example having a strong impact on the air quality).

To evaluate the impact of the i^{th} action at the j^{th} quantum time, i.e. $\mathcal{A}_i^j$, the difference between the following two scenarios need to be calculated: (1) a scenario (the recommended one or any other scenario for that day) ($\mathcal{A}^\star$) and, (2) a second scenario (like a measured one) ($\hat{\mathcal{A}}^j$), the difference is obtained by keeping intact the first scenario except replacing $\mathcal{A}_i^{\star j}$ by the action from the second scenario (like what the occupant has done) ($\tilde{\mathcal{A}}_i^j$) as shown in equation (4.10). Both scenarios are simulated using the physical model of the office. The difference between the effects indicates the impact of not performing the first scenario action $\mathcal{A}_i^{\star j}$.

$$\begin{gathered}\hat{\mathcal{A}}^j = \left\{\mathcal{A}_i^{\star k}; \forall k \neq j\right\} \cup \left\{\tilde{\mathcal{A}}_i^j\right\} \\ \text{where, } \mathcal{A}_i \in \{\zeta_W, \zeta_D\}\end{gathered} \tag{4.10}$$

It is interesting to note that when the differences between these two scenarios are considered, equation (4.11) follows, i.e. the difference in actions is zero for all time slots except the j^{th} time slot and at the k^{th} time slot, and the difference is identical to the difference between actual and the recommended scenario. Hence, by considering change in actions between these two scenarios, the change in the i^{th} action at the j^{th} hour can be isolated and its effect can be investigated.

$$\Delta\mathcal{A}_i^j = \mathcal{A}_i^{j\star} - \hat{\mathcal{A}}_i^j = \begin{cases} 0, & \forall k \neq j \\ \mathcal{A}_i^{\star j} - \tilde{\mathcal{A}}_i^j, & k = j \end{cases} \tag{4.11}$$

Using algorithm 2, the impact of the i^{th} action at the j^{th} hour ($\mathcal{A}_i^j$) can be obtained. For a complete lists of impacts, algorithm 2 has to be repeated for every i^{th} kind of action ($\mathcal{A}_i$) and for every j^{th} time slot. For instance, opening the door between 12 a.m. and 1 p.m. not only impacts the air quality and thermal comfort in the same time slot but also impacts the air quality and thermal comfort in the succeeding time slot (1-2 p.m). This is also an example of a common cause leading to multiple effects.

Algorithm 2 Tabulating differential explanations with contextual causality

Input: $\mathcal{T}_0$: Differential explanations from algorithm 1

Output: $\mathcal{T}$: differential explanations with contextual causalities

From $\mathcal{T}_0$ get: $\tilde{\mathcal{A}}, \mathcal{C} \xrightarrow{\tilde{\mathcal{I}}} \tilde{\mathcal{S}}$
From $\mathcal{T}_0$ get: $\mathcal{A}^{\star}, \mathcal{C} \xrightarrow{\mathcal{I}^{\star}} \mathcal{S}^{\star}$
for $j = t_{start}$ to t_{end} **do**
    for $k = t_{start}$ to t_{end} **do**
        if $k = j$ **then** then $\mathcal{A}^{\star,j} = \tilde{\mathcal{A}}^k$
        end if
        apply differential explanations algorithm 1 to obtain $\mathcal{T}_k$
        Compare between $\mathcal{T}_k$ and $\mathcal{T}_0$ if there is any difference insert an arrow between the between the $\mathcal{A}^{\star,j}$ and the different satisfaction
    end for
end for

4.4.3 Model fragment

The effect variables are caused through intermediate variables like air flow and heat flow.

Using the equations of the energy models, it is possible to generate cause and effect relations between actions and final effects, but the causality between the different levels of the intermediate variables and the final effects are indiscernible because their changes cannot be monitored with the energy models [106], as shown in Figure 4.11.

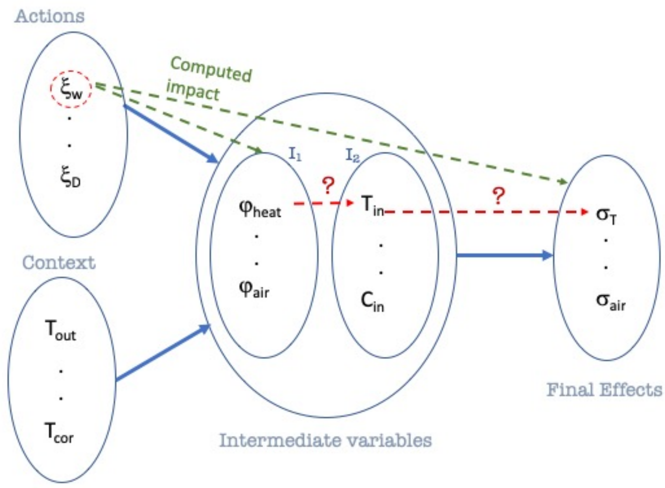


Figure 4.11: Undetected causality between intermediate variables and effects

Integrating relations between the intermediate and the final variables is important to provide occupants with complete explanations. To overcome this difficulty, the model fragment, inspired from GARP3 (presented in the third chapter), is used.

This can be achieved by injecting expert knowledge in the form of model fragments. The model fragments represent potential causalities as well as impossible ones. For instance, heat flow may have an influence on air temperature but not on CO_2 concentration, as shown in Figure 4.12. This can easily be done because the expert has a very good knowledge about the nature of those variables.

Potential causality (pc) is a structural causal relation from a cause variable v1 to a target variable v2. A potential causality does not assume anything about the direction of variation of the values for v1 and v2 (v1 and v2 are labels with domains dom(v1) and dom(v2)).

It is represented as : $v_1 \xrightarrow{pc} v_2$ or $v_1 \nrightarrow^{pc} v_2$ for the forbidden ones.

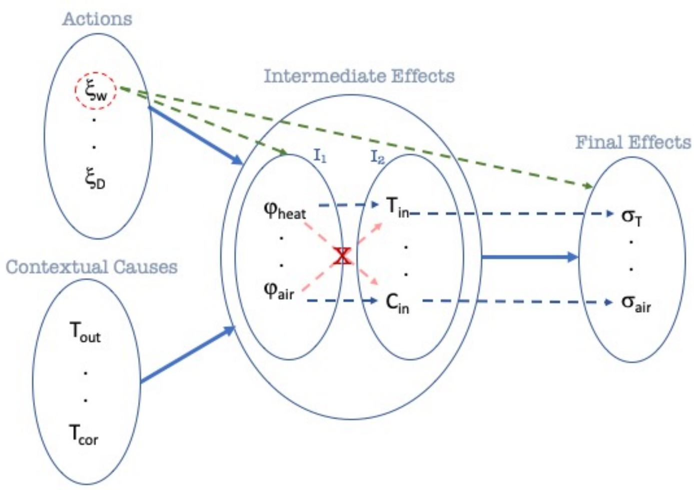


Figure 4.12: The representation of the potential and forbidden causalities

The forbidden causality helps to avoid the false causality caused by the co-occurrence of different events.

Conditional potential causality: A conditional potential causality is activated by a specific condition modelled as a logic proposition applying to values of variables. For example, the causality link between the heater and the inside temperature is correct only when the heater is ON; when the heater is OFF it is a forbidden causality even if a co-occurrence appeared between the heater temperature and the inside temperature.

By integrating calculated causalities and potential ones, a full causal graph for the whole system can be done. Part of this diagram is represented in Figure 4.13 where five categories of nodes appear, viz. actions (red), context (yellow), air flow (blue), heat flow (orange) and effects (green). It can be seen that action nodes have several outward edges and several paths from actions eventually leading to some effects. For example, opening the door between 9 am and 10 am (ζ_D^9) not only leads to thermal comfort (σ_{temp}^9) through heat flow from the corridor (ϕ_n^9) but also leads to air-quality-based comfort between 11 a.m. to 12 p.m. (σ_{air}^{11}).

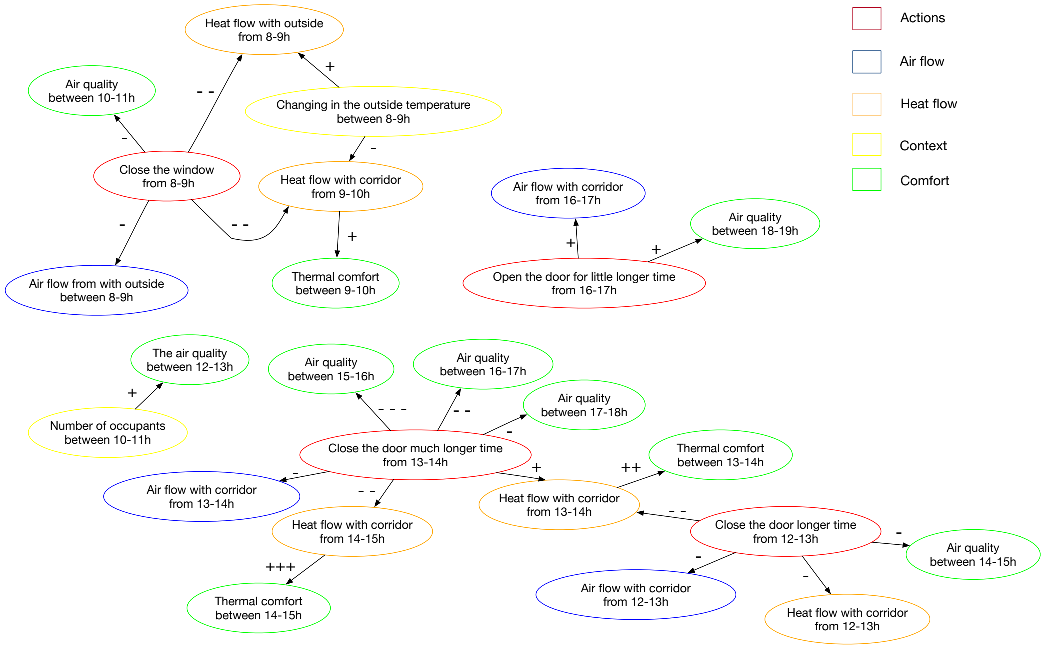


Figure 4.13: Causal Graph

Thus, differential explanations allow the occupants to have an explanation based on the cause-effect relations of their actions and they may decide to change their routines or learn from their historical actions. Chapter six presents how transform the explanations into natural language to be shared with occupants. The next section describes the second form of the causal explanations "direct explanations" based on Bayesian networks. The next section also presents an initial work to present a proof of concept.

4.5 Direct explanations

Direct explanations are different from differential explanations. They do not compare two different scenarios to generate explanations. They are based on the learning of a Bayesian network from many different simulations generated by the EMS for a day. These different simulations with the model fragments (presented earlier) are used as an input for a Bayesian search algorithm to learn the Bayesian network structure.

Bayesian networks are a member of a vast class of models, ones that can be used to describe nested, acyclic statistical models of virtually any kind of non-pathological joint probability distribution [85]. Their signature characteristic is their ability to encode directional relations which can represent cause-effect relationships, compared to other graphical models that cannot, e.g., Markov networks [9].

As mentioned before, the ability to represent directional relationships is an essential reason for choosing to focus on Bayesian networks.

Learning a Bayesian network from data in general involves two sub-tasks: learning the structure of the system (i.e., determining what depends on what), and learning the parameters (i.e., the strength of these dependencies). As it is trivial to learn the parameters for a given structure from a complete data set (the observed frequencies are optimal with respect to the maximum likelihood estimation [17]), for explanations it is more important to focus more on the task of learning the structure.

The Bayesian search structure learning algorithm is one of the earliest algorithms. It was introduced by (Cooper & Herkovitz, 1992) and was refined somewhat by (Heckerman, 1995). It follows essentially a hill climbing procedure (guided by a scoring heuristic) with random restarts.

The Bayesian search algorithm has the following parameters:

- Max Parent Count limits the number of parents that a node can have, because the size of conditional probability tables of a node grows exponentially by the number of the node's parents.

- Iterations sets the number of restarts of the algorithm. Generally, the algorithm is searching through a hyper-exponential search space and its goal can be compared to searching for a needle in a haystack. Restarts allow for probing more areas of the search space and increase the chance of finding a structure that will fit the data better. The number of iterations gives an idea of how long the algorithm will take when the number of iterations is large.

- Sample size is a factor in the score calculation, representing the inertia of the current parameters when introducing new data.
- Seed (default 0), is the initial random number seed used in the random number generator. A seed equal to zero (the default) makes the random number generator really random by starting it with the current value of the processor clock.
- Link Probability (default 0.1) is a parameter used when generating a random starting network at the outset of each of the iterations. It essentially influences the connectivity of the starting network.
- Max Time (seconds) (default 0, which means no time limit) sets a limit on the runtime of the algorithm. It is a good idea to set a limit for any sizable data set so as to have the algorithm terminate within a reasonable amount of time.
- Use Accuracy as Scoring Function (default OFF). When checked, the algorithm will use the classification accuracy as the scoring function in search for the optimal graph.

The algorithm produces an acyclic directed graph that gives the maximum score. The score is proportional to the probability of the data providing the structure, which, assuming that the same prior probability was assigned to any structure, is proportional to the likelihood of the structure given the data. The algorithm allows the injection of expert knowledge in the form of potential/forbidden causalities (model fragments). This is helpful in organizing the variables and eliminating the correlation between variables and extracting causalities.

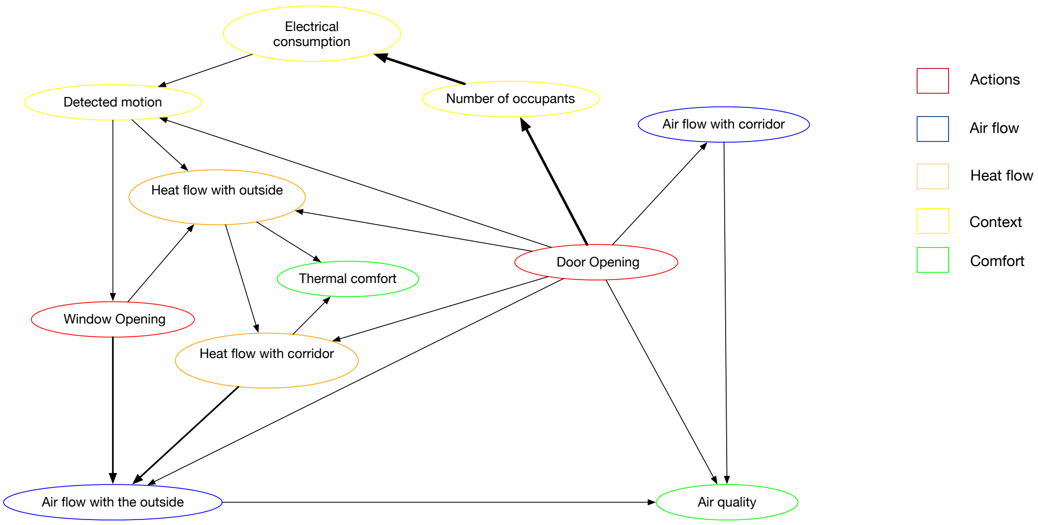


Figure 4.14: Direct explanations

To apply the Bayesian search algorithm in the case study, presented in Figure 4.14, 300 simulations were obtained from a genetic algorithm used by the EMS to optimize actions and find the recommended actions. This might have had an effect on the learning of the structure as the simulations are not completely random. They are oriented by the genetic algorithm that searches the best set of actions according to occupants preferences (more details about the genetic algorithm used are in Appendix 1). The model fragments are presented in the form of potential causalities, as in Figure 4.15, for example the heat flow cannot be the cause for the change in the CO_2 concentration.

Direct explanations present the explanations in the form of probabilistic causality, to represent the different cause-effect relationships between the variables in general.

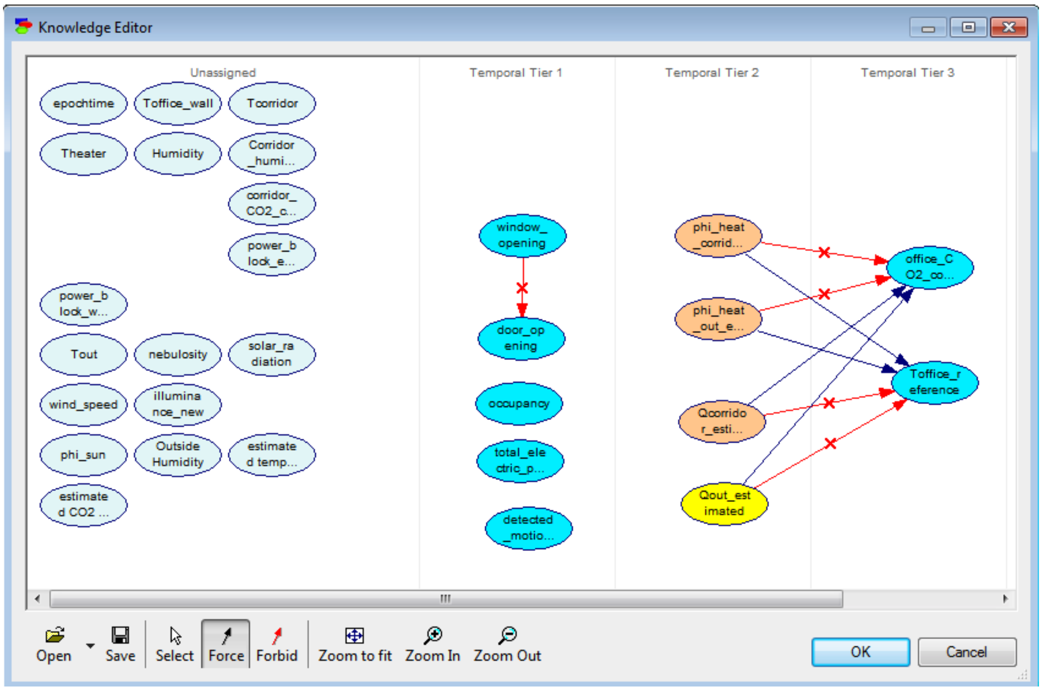


Figure 4.15: Model fragment in the form of potential causalities (figure is realized using the BayesFusion program)

However, direct explanations cannot present the cause-effect relationships for each action at any time. They also cannot represent the delayed impact of each action on the different criteria. Another limitation is that for any one day the algorithm cannot learn the impact of the contextual variables on the comfort criteria because there is not enough variation in the context variables to detect their causality. The inertia in the buildings limits the learning of the Bayesian network, as the impact of the action is delayed by the inertia, and the search algorithm cannot learn that. One possible solution is to use the Dynamic Bayesian network to overcome this difficulty.

This represents a primary test as to what can be done with the Bayesian networks.

4.6 Conclusion

The chapter presents the generation of explanations with the use of knowledge models. Knowledge models can be any type of model that can provide simulation between input variables and output ones, like physical models or linear regression models [105]. This chapter described a real case study with a model-based EMS to illustrate the difficulty in understanding these type of systems, and at the same time why it is very important for occupants to understand the impact of their actions.

Then, it describes the different steps to obtain differential explanations, and how they can help the occupants to understand the impact of their actions. It presents how it is possible to explore the implicit causality in the knowledge based EMS and render it explicit thorough the differential explanations with contextual causality. This chapter also present the model fragments concept to allow the injection of the expert knowledge and help determining the causalities. It also describes direct explanations and their limitations.

As presented in the second chapter, model-based EMSs or knowledge models, are complex and hard to implement. A general model-based EMS that is easy to install and tune is still a continuing field of research [105]. The next chapter will present a new method to enable the generation of explanations when knowledge models are not available.

Chapter 5

Explanation generation without knowledge models

As has been discussed earlier, it is difficult to build an energy model. This is because buildings energy models demand a lot of expert time, important knowledge about the building characteristics, and profound physical understanding. Identified before, when there is a change in the environment, adding an electric heater or other appliances would require the re-calibration, or even modification, of the model. Without a knowledge model, applying the differential explanations method, in the way presented in the fourth chapter, would not be possible, as the proposed method depends on the knowledge model to do different simulations and extract the effect of each action. This is why there is a need to find a new approach to generate explanations without the need for knowledge models. It is necessary to have an approach without the need for an expert to build the energy model. This approach should be easier to construct, scalable and should not need profound physical knowledge. This approach will be based only on the collected data and will present a concept to be able to compare actions and effects.

This chapter proposes a new approach to substitute the knowledge model in building an energy management system with a "data model" to permit the generation of differential and direct explanations, always with the aim of enabling occupants to understand the impact of their actions such as opening/closing of doors/windows and involving them in enhancing their satisfaction without extra cost. This chapter starts by describing the problems lying behind the proposed approach. Then it continues with the steps to overcome those difficulties, and ends with the tests on the experimental testbed and the generation of explanations.

5.1 Contextual statement

As there is a causal relationship between variables in buildings, an action by the occupant, for example, opening the window if the outside temperature is lower than the inside temperature, will decrease the inside temperature, affect the air quality, and might also cause the heater to work for longer with an increase in the energy bill. The occupant's comfort (indoor temperature, CO_2 concentration, ...) in the building is affected by context variables (outside temperature, wind speed,...) and the occupant's actions (opening/ closing windows and doors, turning on heaters, ...). It is possible to present a hypothesis for this; the hypothesis is shown in Figure 5.1 and defined as:

If occupants are on two days having similar context variables, such as similar outside temperature, solar radiation, .. and they perform the same actions, they should obtain similar effects variables and then have similar satisfaction.

Still, there are different challenges to be faced with this approach:

1. How to compare context variables between two days and extract days having similar Context? Which Context features are important to compare two days?

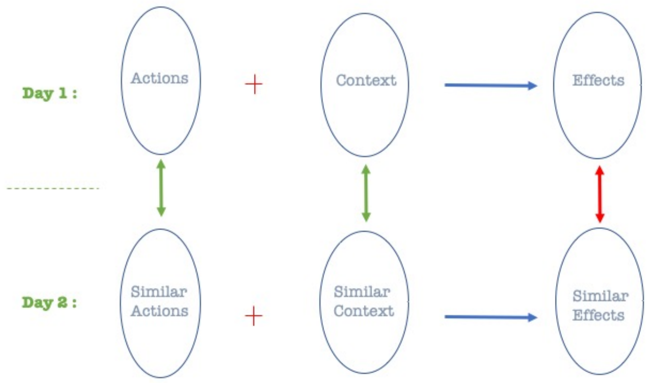


Figure 5.1: Hypothesis for Context, Actions and Effects

2. How to obtain the best set of actions for occupants' satisfaction from the set of days having similar Context?

3. How to generate and validate the recommended actions for occupants?

4. How to use this new "data model" approach to generate the differential/direct explanations?

This approach does not include a knowledge model, therefore there is no way to simulate or calculate the energy scenarios anymore. Besides that, this approach uses only the measured data as a preference for the user and the building compartments within a specific context. This does imply that with the same context doing the same actions should give the same comfort level. For example, opening the window when the outside temperature is higher than the inside temperature will always lead to a gain in heat.

5.1.1 The challenge of extracting similar days

The first and the most difficult challenge in this approach is how to define similar days based on context variables. Different context variables do not have the same importance regarding their impact on the effect variables. For example, the change in wind speed has less impact than the difference between the outside temperature and inside temperature for the case study (this might not be right for other cases).

In the domain of building energy, all of the studies focused on clustering methods to extract similar days. In general, they have used clustering for the selection of typical demand days for the optimal design of building energy systems [108]. They used the input data for the building's heat demand, electricity consumption and the solar irradiation for clustering. They compared k-centers, k-means, k-medians, and k-medoids methods as well as monthly averaged input days.

They have also used a two-level clustering strategy for Energy Performance Evaluation of University Buildings [73]. They aggregated intra-building clustering and inter-building clustering for the final result.

In a different study, they used incremental k-means for clustering in the air population dataset and they used clustering results for the weather forecast [18].

However, these studies only considered the indoor environment or outdoor environment, and they did not take into consideration the different in importance of different variables on the environment. They also did not discuss the importance of occupants' behavior in spite of the impact of occupants' actions on the building.

5.1.2 The challenge of variables weighting

As mentioned, the relative importance of variables/features affects the accuracy of results significantly when using Euclidean distance to verify similar days. In [34] they showed that feature weighting could improve the accuracy of the classical methods. They tried with the K-NN (k-nearest neighbors algorithm) for three databases (Cleveland Heart, hepatobiliary disorders, and Monk problems) and got better results with the weighted data. Hence, it is necessary to determine the importance of each feature; the more the feature is essential, the higher the weight should be.

There are many studies for estimating the importance of features. One study used mutual information for feature selection and feature weighting and applied it to the KNN classification [107]. In another approach, they used a Genetic Algorithm and Information Gain to weight the features and used them in text categorization [72].

The study mentioned that there are different dimensions to be aware of for the features weighting [23]. One critical aspect is the knowledge to constrain the case representation, guide feature transformation, and assign case-specific weight settings.

Differential explanations are based on the comparison between two scenarios or similar ones. This chapter proposes a new approach to identify and extract similar days based on their similar context variables to generate the explanations. To reach this goal different difficulties need to be overcome, like how to compare different variables with different scales and a different nature, then, how to learn their relative importance. Still there is a need to define what does mean "two similar days". The next section presents how this work treats these difficulties, and continues with the different steps to obtain the data model and generate the explanations.

5.2 Proposed solution

The concept of this approach is shown in Figure 5.2. In general, for a specific day, the aim is to find the set of past days having similar context features/variables to this day. After that, from this set of similar days, choosing the day which has the best occupant's satisfaction and obtain the set of actions from this day to recommend to the occupant.

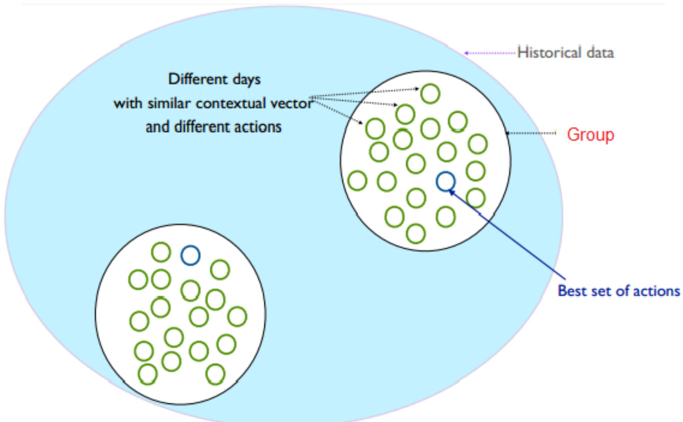


Figure 5.2: Concept of proposed solution

To do this, the solution, shown in Figure 5.3, proposes the following steps:

1. Grouping variables with the same nature and normalize them.
2. Using a **Genetic Algorithm** for feature weighting to determine the importance of Context, Action and Effect.
3. Extracting days having similar Context features
4. Obtaining the best set of actions and recommending them to occupants.
5. Generating the differential and direct explanations.

5.3 Normalizing data in the building

Due to the different physical nature of variables/features in the building, it is necessary to normalize them to have a comparable impact of each feature. In the building area, each feature has a different physical nature and different units to measure them. For example, based on physical knowledge, with CO_2

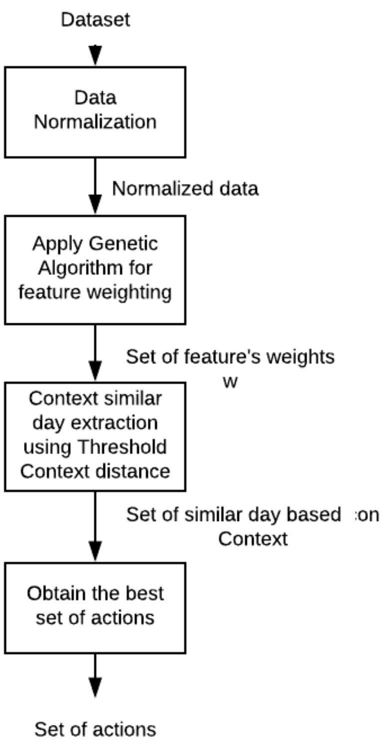


Figure 5.3: General schema to obtain the data model

concentration, it is considered that if the difference is less than (100 ppm particle-per-million), it is a tiny change because the average CO_2 level of concentration from an occupant is about $400ppm$. With the wind's speed, from 0-2m/s, the wind is calm, 2-4m/s is windy. On the other hand, with temperature, if the difference is greater than 2^oC, it is considered a substantial change. This means that the necessary amount of each feature to change the occupant's satisfaction is different.

Hence, it is essential to have a normalization method to balance the values of features in the building.

Granularity based on expert distance.

This work proposes a granularity method based on the physical natures of features: Granularity is based on expert distance. This distance is critical to be able to compare different variables together. The formula for this method is presented as follows:

$$Granularity(x) = \frac{x}{s_x}$$

with s_x - expert distance of features x.
Because of the nature of features in real life, based on the physical nature and expert knowledge, the expert distances of some context features are defined as follows:

- Wind speed: 2m/s (calm:0-2, windy:2-4, stormy:4+ m/s)
- CO_2 concentration: 400ppm (400ppm for a person)
- Temperature: 1 degree
- 0.4% (clear:0-0.4, cloudy:0.4-0.7, covered: 0.7+)
- Laptop power consumption: 17 Watt(W) (if it is > 17W, it means that an occupant is using a PC).

5.4 Similarity method between different days

The primary step of this approach is to define when days are similar and to identify them. It is necessary to establish a method to compare the similarity between two days. After the normalization step, this step defines a similarity-based method to compare two days based on different groups of features Context, Action, and Effect.

1. **Effect similarity-based method.**

   Two days $D1$ and $D2$ are similar if their measured effects are similar. So, the effect similarity-based method is defined using Euclidean distance to determine the similarity of effects between two days. The formula for this method is shown below:

   $$E\_Distance(D1, D2) = \sqrt{\sum_{i=1}^{n}(E_{D1}[i] - E_{D2}[i])^2} \quad (5.1)$$

   with n - the number of sampled Effect features for a day depending on the quantum time for sampling, $E_{D1}[i]$ and $E_{D2}[i]$ - the vectors of the Effect feature at the i time instant of days $D1$ and $D2$ respectively.

   In the case study the occupant's satisfaction is based on thermal comfort and air quality. The aim is to obtain two days having the similar Effect features (inside temperature , CO_2 concentration). To identify whether two days are similarly based on their effect variables, a physical rule is defined .

   Physical rule: Two days D_1 and D_2 are similar if and only if they satisfy the following condition:

   the maximum distance of the difference for each effect variable should not exceed an expert defined level depending on its nature. Within that the condition for the case study is:

   - The maximum difference of indoor temperature is $T \leq 2^oC$
   - The maximum difference of CO_2 concentration is $CO_2 \leq 400ppm$

2. **Context similarity-based method.**

   The Context similarity-based method is defined to compare context features between two days $D1, D2$ as the following:

   $$C\_Distance(D1, D2) = \sqrt{\sum_{i=1}^{n}(w_C[i] * C_{D1}[i] - w_C[i] * C_{D2}[i])^2} \quad (5.2)$$

with $w_C[i]$ - the weight of context feature i (it represents the relative importance of the feature), n - number of context feature, $C_{D1}[i]$ and $C_{D2}[i]$ are the context features i of day $D1$ and $D2$ respectively.

3. **Context, Action similarity-based method.**

   Similarly, as the combination of context and action, it defines the distance of two days based on both context and actions as:

$$CA\_Distance(D1, D2) = (\sum_{i=1}^{n_C}[w_C[i] * (C_{D1}[i] - C_{D2}[i])]^2 + \sum_{i=1}^{n_A}[w_A[i] * (A_{D1}[i] - A_{D2}[i])]^2)^{\frac{1}{2}}$$

   with $w_C[i], w_A[i]$ - the weight of context, action feature i, n_C, n_A - number of context, action features, $C_{D1}[i]$ and $A_{D1}[i]$ are the context, action features i of day $D1$ respectively.

From these similarity-based methods, the similarity between two days is determined by the distance between them. However, in all of the above similarity-based methods, the weights of context, action features always play an important role, so it is essential to determine the best weights for context, action and features.

The global objective is to find the best set of weights for context and action features w_C, w_A, which satisfies the condition: if a context or action feature is more important for effect features, it should have a higher weight and if two days have a small CA_Distance, they should have a small E_Distance and vice versa. It is important to combine the context and actions features. For example, the wind speed importance might vary depending on the opening/closing of the window. This is why features weighting needs to combine the context, action variables for better results.

There are some feature weighting tools such as information gain, mutual information, Laplace score, but here, where there are many features with the complexity of the objective, it is necessary to have an approach for estimating the weights of many features and satisfying defined conditions. Hence, the genetic algorithm, an evolutionary search algorithm proposed as a multi-objective optimization method to use for feature weighting.

Genetic Algorithm (GA) for feature weighting

As mentioned, the objective with the genetic algorithm is to find the best set of weights for context and action features w_C, w_A, which satisfies the condition: the relative importance of a context or action feature is higher or lower depending on its impact on the effect features.

$$
\begin{aligned}
w_C, w_A : \text{minimize} \quad & E\_Distance(e_i, e_j) \\
s.t \quad e_i, e_j : & CA\_Distance(e_i, e_j) - small
\end{aligned}
$$

Where e_i, e_j candidates of solution generated by the genetic algorithm (chromosome).

GA tries to minimize E_Distance within a group of days having similar context and action features.

To deal with the mentioned objective, each chromosome's (candidate of solution), CA_Distance is used to group days having similar context and action to the same group. Then the algorithm tries to minimize the E_Distance within each group. This results in days having a small *CA_Distance* being placed in the same group. So, if the E_Distance within every two members of the group is small, it means that members of the same group have a similar effects. This would satisfy the objective: CA_Distance is small ⇒ E_Distance is small. The task now is minimizing the E_Distance within each group.

The number of members within each group is related to the diversity of the set of days having similar context and action. Another objective is to minimize the number of groups to maximize the number of members within each group.

With these objectives, the approach is to estimate the fitness score achieved with the following steps.

1. **Clustering data based on CA_Distance**.
   With each chromosome as a solution to w_C, w_A, k-means clustering applied with the number of cluster **m** to group the data, the distance method is CA_Distance. The objective is to cluster days having a small CA_Distance in the same group.

   We defined C_{w_C,w_A} as the clusters obtained by grouping with w_C, w_A.

2. **Fitness score formula to minimize E_Distance**.
   After the clustering step, the days which are in the same cluster will have a similar CA_Distance.

   The objective is to minimize the E_Distance between days having similar Context and Action. So, with a fitness score for days e_i, e_j in the same cluster C_i, the aim is to minimize the E_Distance between days in the same group. Moreover, it is necessary to minimize the numbers of clusters to maximize the diversity of members in each cluster, to find and maximize the opportunity of finding enough diversity to recommend actions. For that, a compromise is needed between the CA_Distance and the number of clusters.

   Hence, the algorithm tries to minimize the fitness score with candidates

w_C, w_A:

$$D_{w_C,w_A} = \frac{1}{m} \sum_{C_i \in C_{w_C,w_A}} \left(\sum_{e_i,e_j \in C_i} \frac{E\_Distance(e_i, e_j)}{|C_i|} \right) \tag{5.3}$$

with e_i, e_j - cluster members , m - number of clusters, $|C_i|$ number of members in the cluster i.

The desired solution is defined as:

$$w_C^*, w_A^* = argmin \quad D_{w_C,w_A} \tag{5.4}$$

From the formula, it can be observed that if the days in the same cluster have a similar Effect, the E_Distance is small and it leads to the fact that D is small and that Fitness is high (because it is looking for the minimum D).

This step leads to the definition of weights in distances.

5.5 Extract similar days based on Context

After having found the set of context, action features weights w_C and w_A, the next objective addressed is to use them to extract similar days.

When applying this approach in real life, the set of actions of a new day is unknown, it leads to the consequence that the Effect of this day is also unknown, the only available information is Context features (using the weather forecast). Hence, when extracting similar days to obtain the best set of actions, only context features and C_Distance are used to compare the similarity between days.

For a specific day D, it is possible to extract similar Context days of D, though, there is no fixed number for similar Context days. If the number of similar Context days is too big, there might be several non-similar days in the cluster. On the other hand, if there is a small number of similar Context days, the diversity of the set of actions may not be enough to be able to recommend actions for occupants.

As defined in the physical rule, two days are similar in effects if the difference of temperature is less than 2^oC and the difference of CO_2 concentration is less than $400ppm$.

So, the objective in this part is to extract similar days based on C_Distance and satisfy the physical rule of Effect.

Threshold context distance

Here, the objective is to find the set of similar Context days of a day D that satisfies the condition: all days in this set are similar to a day D based on the physical rule and context features. With this condition, when applying the set of actions of each day D_i in this set to a day D, the effect obtained in the day D will be similar to the effect of D_i.

Hence, this approach defines T_C - a threshold context distance for two days $D1, D2$ fulfilling the condition that if the context distance between them is less than T_C, they will be similar based on the physical rule. In other words if the occupant applies the same actions for all the days that satisfy this distance, the difference on the effect variables will not be sensed by the occupant (depending on the physical rule).

$$C\_Distance(D1, D2) \leq T_c \Rightarrow \text{D1, D2 - similarly based on physical rule} \tag{5.5}$$

However, it is difficult to determine precisely the T_C satisfying the above condition. If T_C is too small, it is not always possible to obtain some similar days satisfying the condition (recall). But if T_C is too big, it might have some incorrect days in the result (precision). The example of this challenge is shown in Figure 5.4.

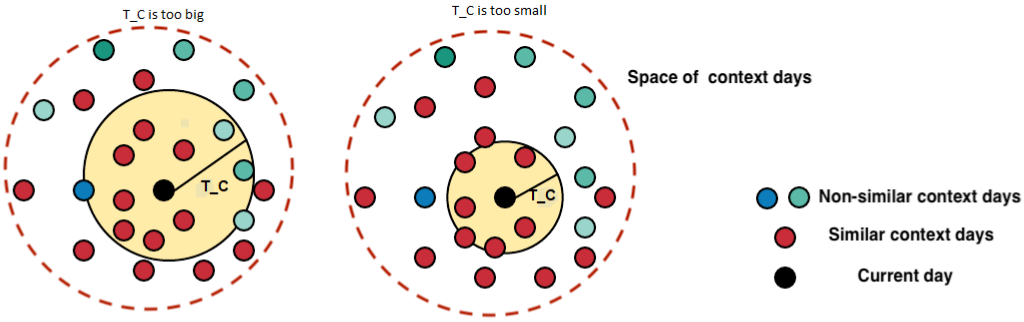


Figure 5.4: The challenge of threshold Context distance T_C

To solve this problem, a classification method with the F1_score is proposed to determine the T_C, because the F1_score is related to both precision and recall.

The F1_score is defined as follows:

$$\begin{aligned} F1\_score &= \frac{2}{Precision + Recall} \\ Precision &= \frac{TruePositive(TP)}{TruePositive(TP) + FalsePositive(FP)} \\ Recall &= \frac{TP}{TP + FalseNegative(NP)} \end{aligned}$$

Specifically, each pair of days is labelled based on both the physical rule and T_C. L is defined as the set of labels for similar days based on the physical rule of Effect. Each label of L is determined as following:

$$\forall l_i \in L, l_i(D1, D2) = \begin{cases} 1 \text{ if D1,D2 - similar based on physical rule} \\ 0 \text{ else} \end{cases} \tag{5.6}$$

T_C is determined such as:

$$T_C = argmax(F1\_score(L))$$

This means that the chosen T_C will ensure the best compromise between the accuracy and obtaining the maximum number of similar days based on their effects for a day D.

5.6 Extract the best set of actions

After having the set of similar Context days in the previous step, the objective now is to extract the best set of actions to recommend to occupants for enhancing their satisfaction.

As the aim is to strengthen the occupants' comfort, the plan is to find the day having the best satisfaction from the set of similar Context days. Then, extract the set of actions from this day to recommend it to the occupants. The concept and main steps of the approach are exhibited in figure 5.5.

From the set of similar days based on Context **SC**, the best day D^* to get the set of actions is defined as following:

$$D^* = argmin_{D \in SC} S(D) \tag{5.7}$$

where $S(D)$ - satisfaction (comfort) of day D based on the effects variable.

Next, we will focus on the real tests in the case study and the experimental results with the proposed solution.

The same testbed from the last chapter was re-used. Table 5.1 describes the different variables and their units in the testbed.

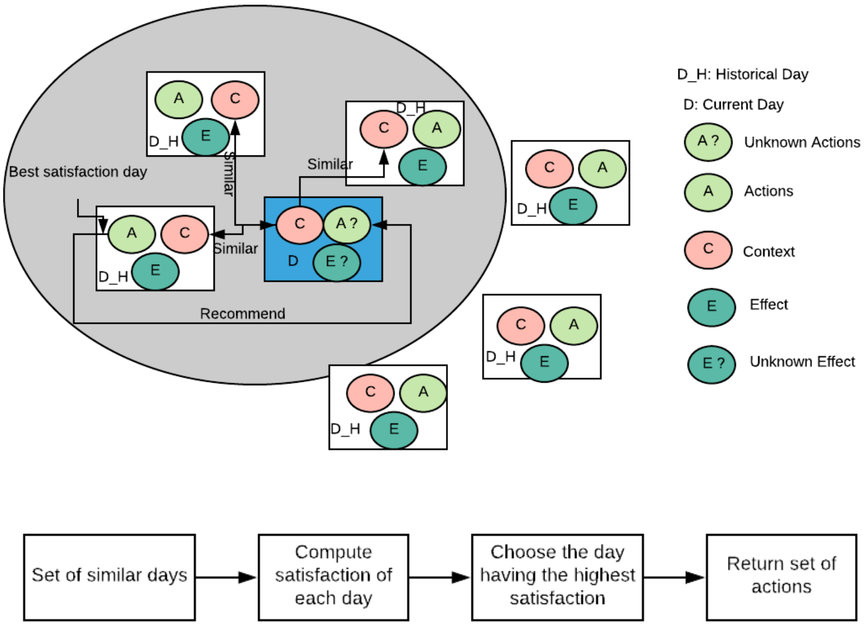


Figure 5.5: Obtain the best set of actions

5.7 Dataset

The sensor data and weather conditions were recorded for a consecutive period of about one and a half years (April 2015 to October 2016) at hourly intervals and stored in the historical database for future simulations (using physical models), so that occupants can learn from past mistakes.

During the summer (from April to September), the heater does not work, and all sensors are disabled during some holidays such as 29/05/2015 - 01/06/2015 and 20/06/2016 - 22/06/2016.

Only working days were chosen (no weekends or closed days,as there are no occupants in these days) for periods: **01/05/2015-28/05/2015**, **01/06/2015-23/07/2015**, **11/05/2016 - 31/05/2016** and **02/06/2016 - 19/06/2016**; after filtering, **100** days are found for training the model. For

Feature	Unit	Description	Type
Toffice_reference	°C	Indoor temperature	Effect
office_CO2_concentration	ppm	Indoor CO2 concentration	Effect
Tcorridor	°C	Corridor temperature	Context
Illuminance (new,old)		Luminosite in office	Context
solar radiation	W/m2	Solar radiation in office	Context
wind speed	m/s	Outside wind's speed	Context
corridor_CO2_concentration	ppm	Corridor CO2 concentration	Context
power (block east, west, total)		electricity consumption	Context
occupacy		Number of occupacy in office	Context
nebulosity	%	Outside nebulosity	Context
Tout	°C	Outside temperature	Context
window_opening	minute	The duration of opening window	Action
door_opening	minute	The duration of opening door	Action

Table 5.1: Table of features in the building

the testing, Another period is selected: **22/06/2016 - 30/07/2016** and has **34** days. The test periods were in summer, so the heater during these periods was turned off.

5.8 Validation and results

5.8.1 Knowledge model for validation

As the problem is related to the occupant's behavior, there is not a standard baseline to validate the results. So, the energy model of the office (presented in the previous chapter) was used. The model shown in Figure 5.6 allows the simulation of the environment and estimation of the occupant's satisfaction. To be consistent with the results, all real Effect features from the registered data, are replaced by the simulated values based on real Context features and Action features.

For a specific day D having Context C, suppose $D1$ as the best historical day we selected, $A1$ is the best set of actions obtained from $D1$ to recommend

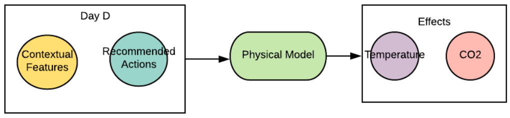


Figure 5.6: Simulation model for the building

to occupants and E^* is the measured Effect variables of the building when applying a set of actions $A1$ to the context C. E^* is simulated from C and $A1$ using a simulation model.

$$C, A1 \xrightarrow{Simulation} E^*$$

Here the aim is to validate the results for two objectives: validate the data model and validate that the recommended actions will enhance the occupant's comfort.

5.8.2 Validating the data model

This step aims to validate the correctness of the hypothesis about the relations between Context, Action, and Effect. Since $D1$ is the best historical day we selected, D and $D1$ which have a similar Context, then the effects E^* should be similar to the original effects $E1$ of the day $D1$. This condition was used to validate the hypothesis. The concept of this step is shown in Figure 5.7.

Figure 5.9 shows some results of the applied experiment. As it can be seen, when applying the same set of actions to two days having similar Context, their Effects are similar, too. For example, with **09/07/2016**, the proposed approach finds the best historical day is **21/07/2015**. With these two days, when applying the set of actions from **21/07/2015** to **09/07/2016**, the indoor temperature and CO_2 concentration we obtained are similar to those in **21/07/2015**, the difference of indoor temperature between the two days is about 1.5^oC and this metric for the CO_2 level is less than $150ppm$.

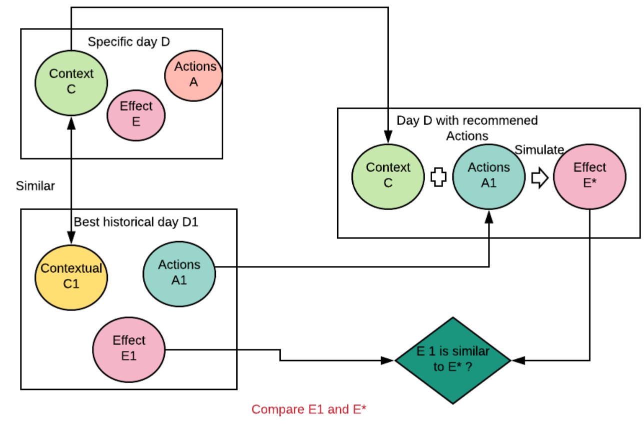


Figure 5.7: The concept of hypothesis validation

ΔT - The difference of indoor temperature (^oC)

	Mean value	Max value
Mean difference	0.59	1.26
Max difference	1.56	2

$\Delta CO2$ - The difference of indoor CO2 concentration (ppm)

	Mean value	Max value
Mean difference	55	139
Max difference	199	400

Figure 5.8: Table of results for validating the hypothesis

With each day D of **34** testing days, suppose that $\Delta E = |E^* - \bar{E}|$ and $\Delta T, \Delta CO2$ are the difference of indoor temperature and CO_2 concentration of ΔE respectively. ΔT and $\Delta CO2$ are computed by the maximum difference between each hour of two days respectively. It has been tested with **34** days, and the result is showed in Figure 5.8. The mean difference of temperature - ΔT is about 1.56^oC and the mean difference of CO_2 - $\Delta CO2$ is about $199ppm$ with all pairs of days. The maximum value of ΔT is 2^oC and the maximum

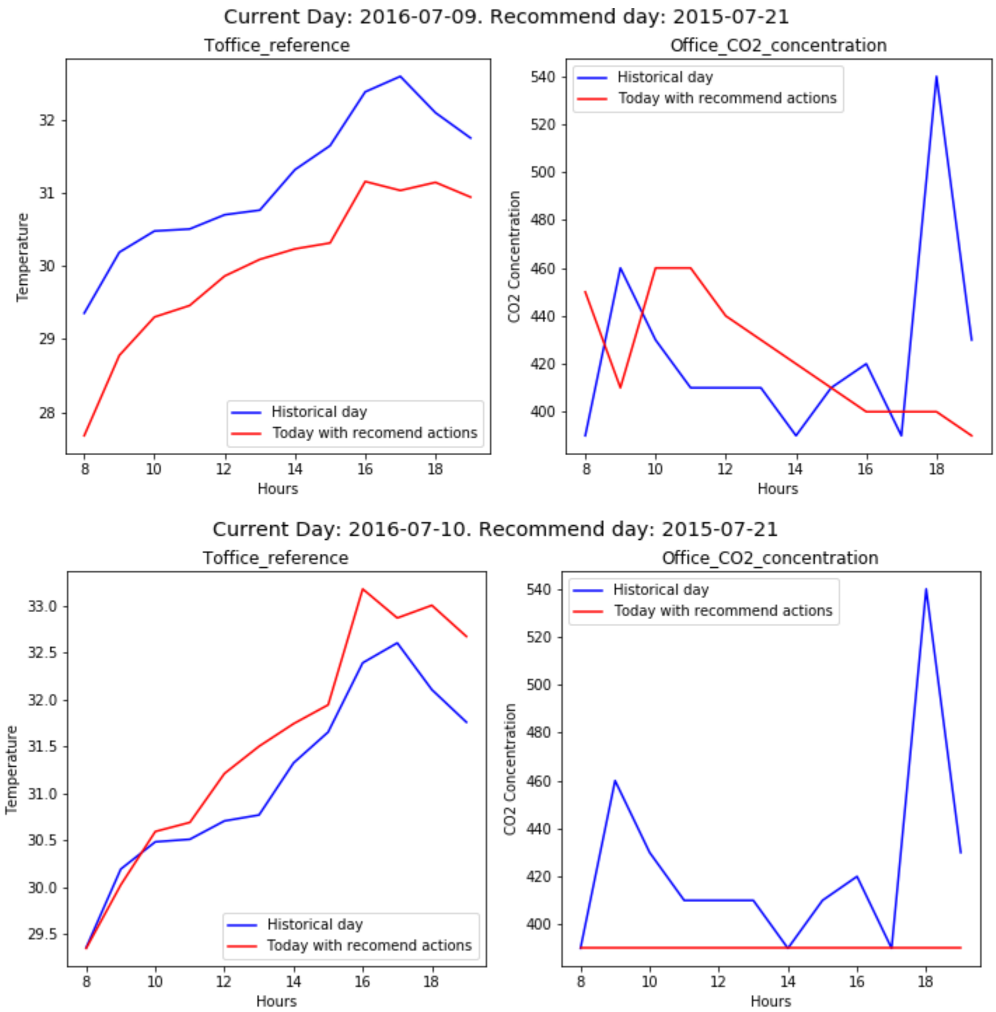


Figure 5.9: Results of applying recommended actions

value of $\Delta CO2$ is $400ppm$.

It demonstrated that when applying the same set of actions to the days having similar Context, they would obtain a similar Effect and the experimental results proved the hypothesis.

5.8.3 Validating the recommended set of actions

This step aims to verify that the recommended set of actions would enhance the occupant's satisfaction. To do that, it tries to compare the Effect E^* of the day D when using recommended actions A^* to the real Effect E with the real actions A of occupants on this day. From that, it could estimate the enhancement of recommended actions to the occupant's satisfaction. The concept of this validation method shown in Figure 5.10.

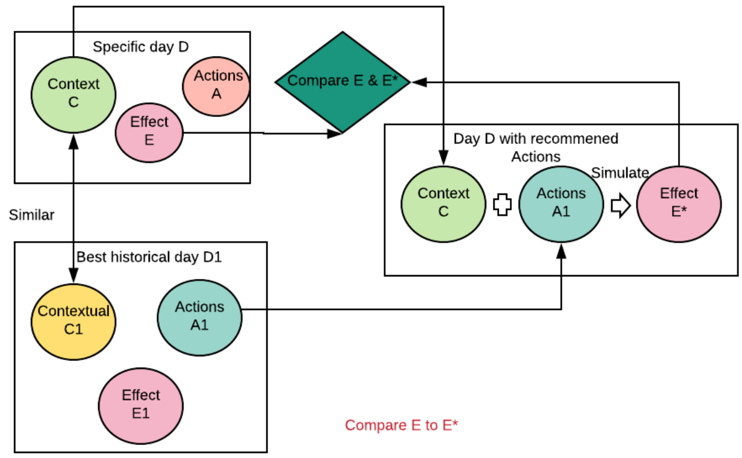


Figure 5.10: The concept of recommended actions validation

For this validation test the occupant satisfaction S is composed of thermal and air quality satisfaction. They are defined as follows:

- Thermal satisfaction: This considers that $T(t)$ gets the highest value (best thermal comfort) when the temperature is between $\mathbf{21}^o$C to $\mathbf{23}^o$C, and the thermal comfort is considered good if $\mathbf{T(t)} \geq 0$ and the good temperature is from **18**^{o}C to **26**^{o}C.

$$TS(t) = \begin{cases} \frac{T-18}{21-18}, & \text{if } T < 21 \\ \frac{26-T}{26-23}, & \text{if } T > 23 \\ 1, & \text{if } T \in [21, 23]. \end{cases} \tag{5.8}$$

- Air quality satisfaction: The air quality satisfaction is calculated using the formula:

$$CS(c) = Min(1, \frac{500 - c}{1500 - 500}) \tag{5.9}$$

Within that the global satisfaction S is :

$$S(t, c) = \alpha * TS(t) + (1 - \alpha) * CS(c) \tag{5.10}$$

with $\alpha \in (0, 1)$ - it represents the relative importance of thermal satisfaction to air quality satisfaction.

Suppose that S^* is the occupant's satisfaction obtained by the Effect E^*, and S is the occupant's satisfaction derived by the Effect E. $D(S, S^*)$ is defined as the day D with real satisfaction S and recommended satisfaction S^*. So, the enhancement obtained when applying the recommended set of actions is computed as follows:

$$H(S, S^*) = \frac{S^* - S}{abs(S)} \tag{5.11}$$

with abs(S) - absolute value of S.

If $H(S, S^*) > 0$, it means that the recommended actions enhanced the satisfaction of occupants compared to the real actions.

With B - the set of days $D(S, S^*)$ with real satisfaction S and recommended satisfaction S^*, the accuracy of the proposed solution is estimated as follows:

$$\begin{cases} B^+ = \{D(S, S^*) \in B : H(S, S^*) > 0\} \\ Accuracy(B) = \frac{|B^+|}{|B|} \end{cases} \tag{5.12}$$

Where B^+ is the group of days where the recommended actions enhanced the occupants comfort.

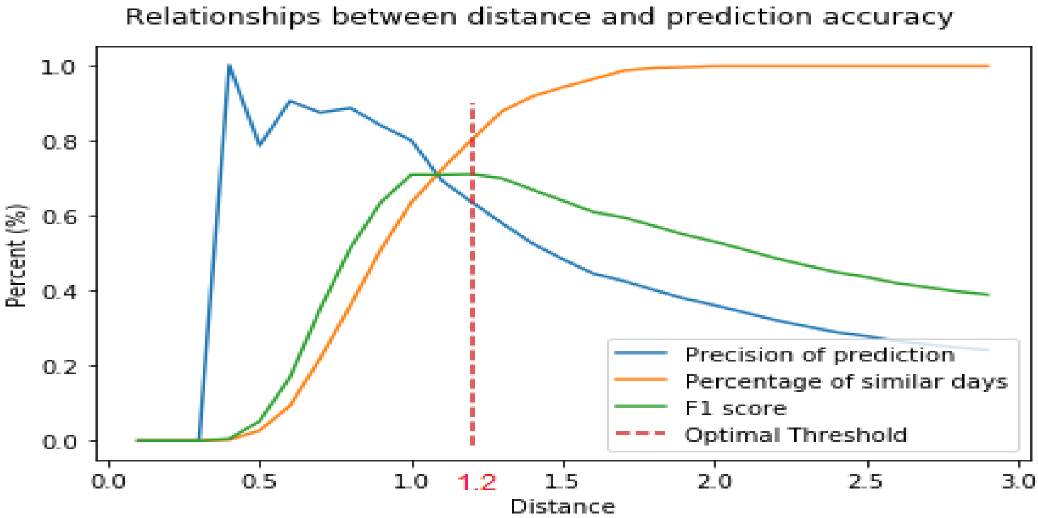


Figure 5.11: Result of threshold Context distanceT_C approach

The result of the threshold Context distance is shown in Figure 5.11. With the case study of this report, the threshold distance $T_C = 1.2$. The results showed that if the threshold distance T_C increases, the number of similar days that could be identified will increase, but on the other hand, the accuracy decreases in this case. If $T_C < 0.5$, the precision (accuracy of similar day prediction) is very high (>85%) but the recall of predicting (percentage of similar days) is low (10%). On the other hand, if $T_C > 1.5$, the precision is lower (50%) but the recall is higher (80%).

However, due to the limitation of real data being **34** testing days, we could recommend a set of actions to only **13** days, so there are **21** days we could not find to similar days with $T_C = 1.2$ from the past days. This is the result of comparing recommended satisfaction with real satisfaction in this case.

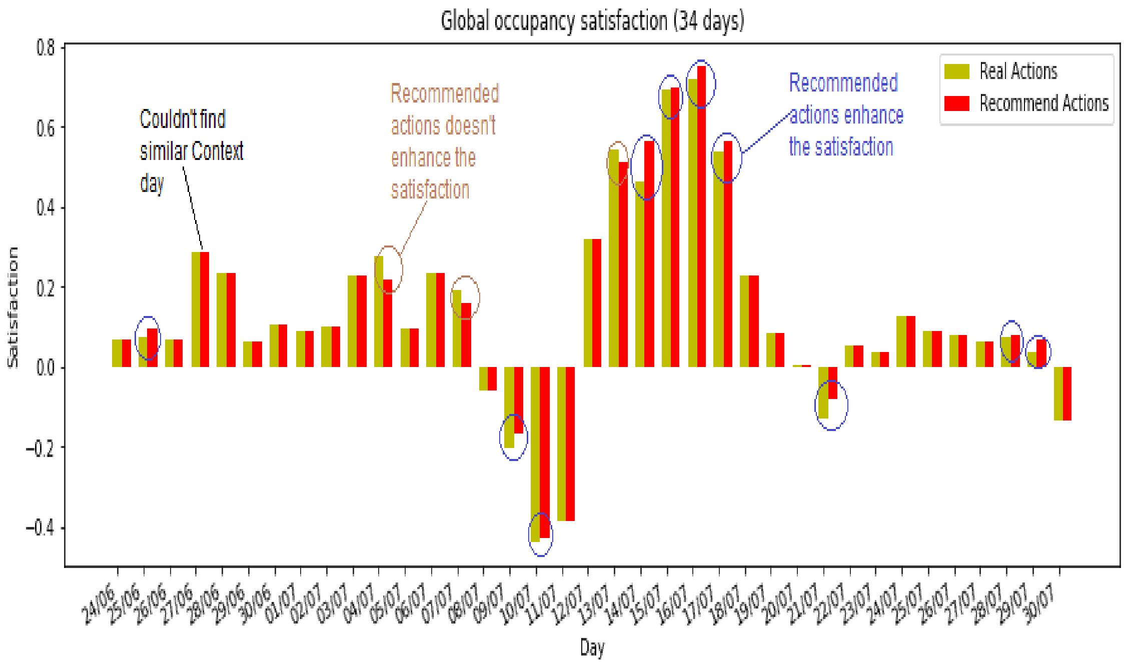


Figure 5.12: Results of recommended actions to enhance global satisfaction (When no similar context day is not found the same actions of the test day are used for simulation)

From the Figure 5.12, it can be seen that, with **13** days, the recommended set of actions could enhance the satisfaction for **10** days (77%). There are **21** days where the approach could not find similar Context days based on T_C and recommend actions due to the limitation of data.

5.9 Analysis and challenges

In different days, the recommended actions could improve from 8% to 18% occupants' global satisfaction. Regarding the example shown in Figure 5.13, when applying recommended actions to **14/07/2016** and **16/07/2016**, the global satisfaction was better by 18% and 11% respectively than real satisfaction; the detailed results of these days are show in Figure 5.13.

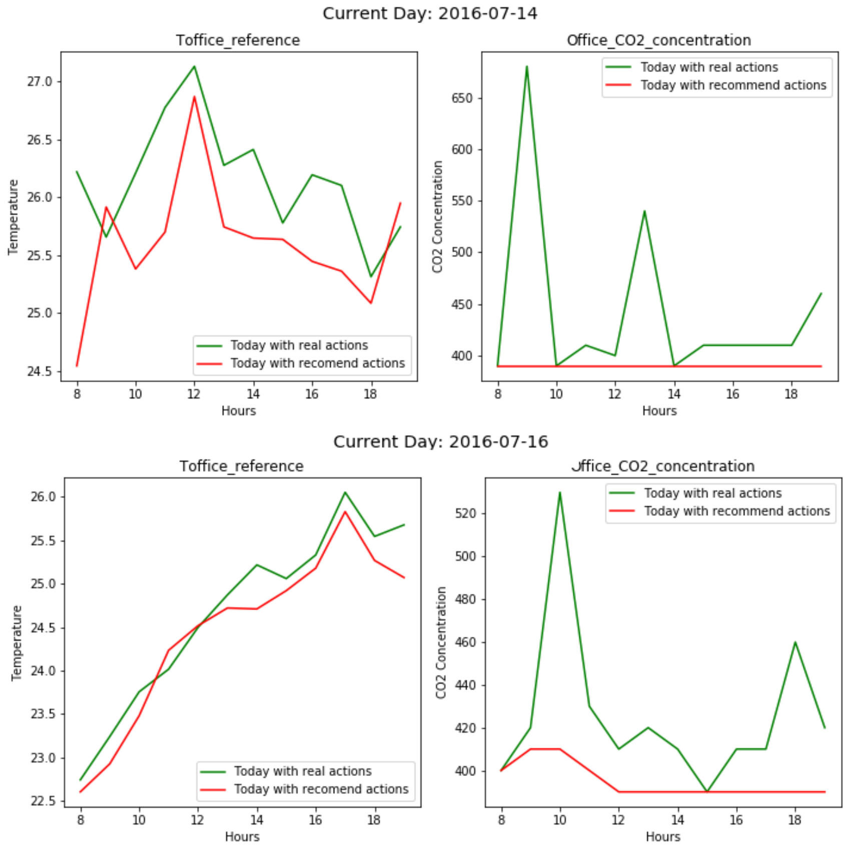


Figure 5.13: Results of applying recommended actions to 14/07/2016 and 16/07/2016

It can be seen in Figure 5.13, for **14/07/2016**, when applying recommended actions, the recommended actions could reduce the indoor temperature by about $1^{o}C$ and with **16/07/2016**, the indoor temperature could decrease by $0.5^{o}C$.

Limitation of data is the challenge of this approach. For some days, the work could not find any day having similar context features because of the restriction of data. It is necessary to have enough significant data to find similar days and choose the best set of actions.

Nevertheless, the data model concept is validated and similar Context days have been obtained, so it is possible now to apply the generation of explanations to the data model. The next section presents the generation of the differential explanations. Then it continues with the generation of the direct explanations.

5.10 Generating differential explanations from the data model

As mentioned and validated earlier in this chapter, when having the same context, it is possible to compare directly between actions and effects. Based on this principle it is possible to use the concept of differential explanations, presented in the previous chapter, on similar Context days. As there is no knowledge model, there is no way to simulate the different scenarios for the same day. Instead here the differential explanations are going to be applied on similar Context days. So the comparison is done between different days with similar context. The difference in user actions is the only cause of the differences in the effects. Applying the differential explanation concept on similar days would generate the descriptive differential explanations as in Figure 5.14. It can only describe the difference in the registered data (actions and effects). In this example, the comparison is done between a random day and the best similar day (the similar day with the highest user satisfaction) and the proposed actions for that day were presented to the user as recommended actions to carry out in order to enhance his comfort.

Using the data model, it can be seen that it is impossible to simulate the impact of each individual action and so it is difficult to estimate the contextual causality of each action. The descriptive differential explanation

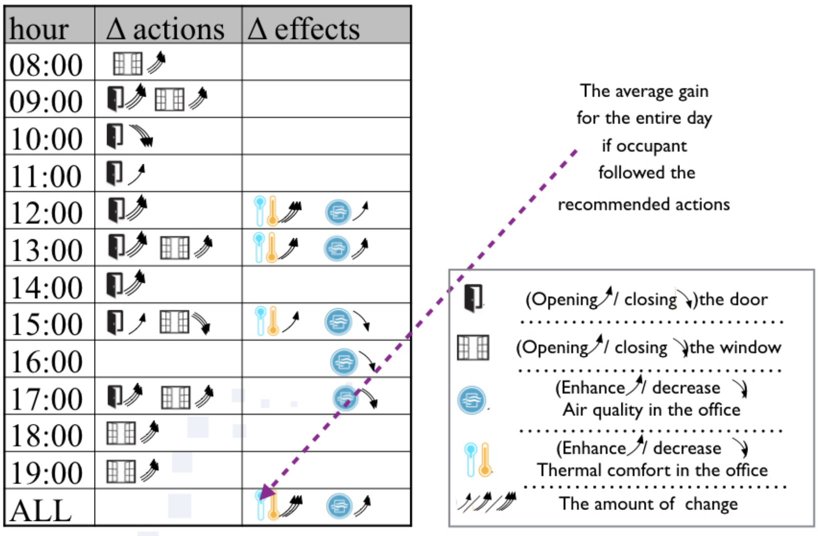


Figure 5.14: Descriptive differential explanations

will generate only the difference between the registered actions and registered effects. It is only based on the observation without any control.

The next section describes the generation of the direct explanations using the data model.

5.11 Generate Direct explanations from the data model

The thermal and air dynamics inside the office on similar Context days are similar as validated before. The effect of actions and behavior of the occupants and the building are similar to each other on similar days. Following from that, learning a Bayesian network to extract the relations between variables on similar days is easier, and the probability of succeeding is higher than when trying to learn a general Bayesian network for all different days. This is justified as there are contradictory behaviors on different days. For that reason, the Bayesian structure is learned on similar days, using the algorithm presented in the previous chapter, to retrieve the correlation between variables and after that tries to extract causalities among them.

Figure 5.15 shows the result of applying the Bayesian search algorithm on the similar days found using the data model on the testbed. This example is built on 8 similar Context days identified by the data model.

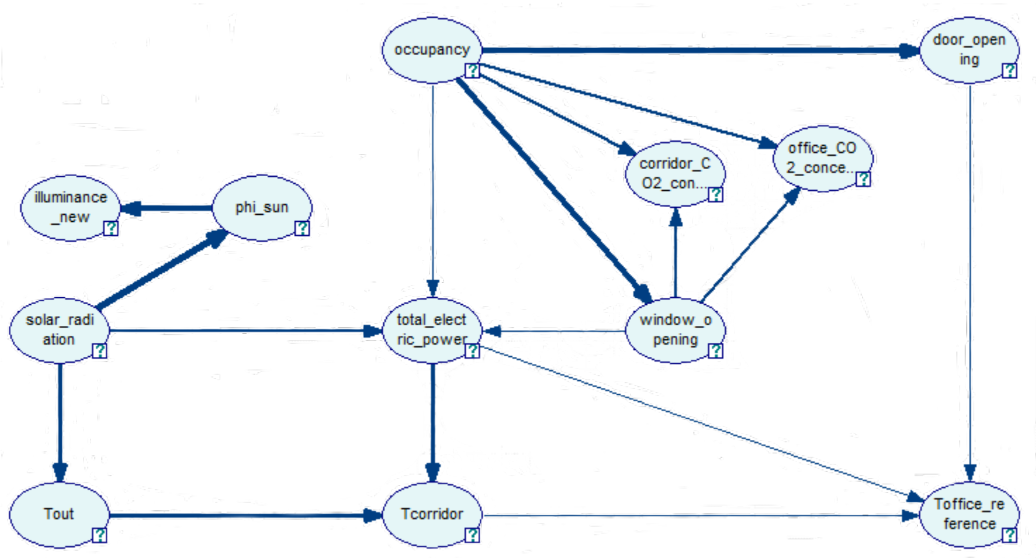


Figure 5.15: Bayesian network for similar days

It can be seen that the Bayesian search algorithm learns the relation between different variables using the data model better than by using the knowledge model. This is the result of learning the network structure using the knowledge model, when there is not a lot of variation in the context variables. This is why the algorithm cannot determine the strength of those links.

5.12 Conclusion

As stated earlier, designing and implementing a knowledge model is not easy and is still an ongoing research subject [105]. For that reason the concept of the data model was developed in this chapter to allow the generation of explanations without the need for a knowledge model and depending only on the occupants' measured data. In this way the generation of explanations can be deployed in almost any type of environment.

	With I/O models	Without I/O models
Differential explanations	✓	✓
Direct explanation	✓	✓

The next chapter describes the form of the explanations and how they are going to be integrated into the e-consultant (the goal of the INVOLVED project). It then validates the work with a field study.

Chapter 6

First user evaluation

The form of explanations is very important to facilitate the cooperation between occupant and the EMS. This work chose human natural language as the form to present the differential and direct explanations, as people use it and find it a very intuitive way to communicate. Then it presents how the explanations are going to be integrated into the buildings' EMSs through a human-machine interface. All of this is done with the aim of building a cooperative system (e-consultant) between the occupants and the EMS.

Although different studies have stated that explanations are important and beneficial for users, as presented in the first and the second chapters, this chapter confirms those studies with a field study with different people to validate whether they would appreciate causal explanations with the EMS or not.

6.1 Text Generation

In the INVOLVED project the main objective is to develop a persuasive system, to engage occupants by helping them to understand and adapt their building EMS to their own comfort criteria. While explanations can help occupants understand how their habitat is working, it is important to deliver them in a language that is the most clear, natural and understandable by different occupants from different ages and different backgrounds. Therefore, this study chose natural spoken language to communicate with occupants. To continue in the same stream of helping occupants, a persuasive interactive human-machine interface specially constructed with the aim of guiding occupants to understand, command and adopt their EMS (this will be explained in more detail later in the chapter) is presented. This part of the work was done with the assistance of the GETALP team of the LIG laboratory, specialized in the natural language processing, where the system to generate natural language "Ariane-Heloise" was developed.

6.1.1 Objective

The differential and direct explanations are provided to help inhabitants to better understand their home environment because they will easily understand the ecological impact of their behavior and how they can modify it to improve that impact. Still, reading a complex graphic table may be a rather abstruse experience for users, and incorporating their knowledge in such a table as the ones presented in the fourth chapter, is not desirable; occupants will favor a statement in natural language rather than the table or a graph. The statement needs to be well written and also entirely automatically generated in a smart way so as not to repeat the same words all the time. Thus, by making a cooperative effort with occupants, it will be easier for them and they can more quickly adopt the right behaviors and enjoy the satisfaction that comes from understanding the effects and their behaviors.

6.1.2 Problem statement

Natural Language Generation (NLG) is an NLP (natural language processing) sub-field concerned with the generation of texts from non-linguistic data [97]. Typical systems take a high-level representation of input data, select and structure the final information to convey (macro-planning), make a lexical and syntactic transformation of this structured information (micro-planning) and finally linearize it as text (surface generation). This classical pipeline has mainly been approached by expert and grammatical rules [94], statistical models [79] or machine learning [112] in applications as diverse as complex medical data summarization [94], prose generation [79] and image captioning [112].

Natural language g (NLG) is a dialog system and a Machine Translation (MT) system. Most recent stochastic MT systems perform natural language generation through shallow language models but MT systems use a high-level internal representation that includes a deep general-purpose NLG module.

MT systems have to generate a target text that complies with the grammar rules of the target language, starting from a deep formal linguistic structure obtained from the linguistic analysis of a text in the source language. Ariane-Heloise was chosen to be used as the MT system to be able to quickly generate the text form of the differential explanations.

6.1.3 Ariane-Heloise system

The Ariane-Heloise MT system [11] is a re-engineering of the widely known Ariane MT system [13]. It takes over all the functionalities by improving or simplifying them. As well as its predecessor, it links together several transducers, which enables it to transform the source text into a decorated tree and to transform this tree into a linguistic tree structure, and then to generate from that structure a target text. During the process, various linguistic operations are carried out: morphological and structural analysis, lexical transfer, syntactic and morphological generation. The decoration on each node of the decorated tree is a combination of values for a set of declared variables and allows for coding of linguistic properties and relations. This part of the work has been done by Jean-Philippe GUILBAUD [1].

6.1.4 GRA-FRA

A first model was developed in Ariane-Heloise for a feasibility demonstration: the GRA-FRA ("GRAphe vers texte en FRAnçais": graph to French text) model. It was specifically designed for generating messages from the tables containing differential explanations. It receives input text that is the linearized arborescent expression of the tables and it produces a French text. The 7 phases of contents processing are succinctly described below from a simple example.

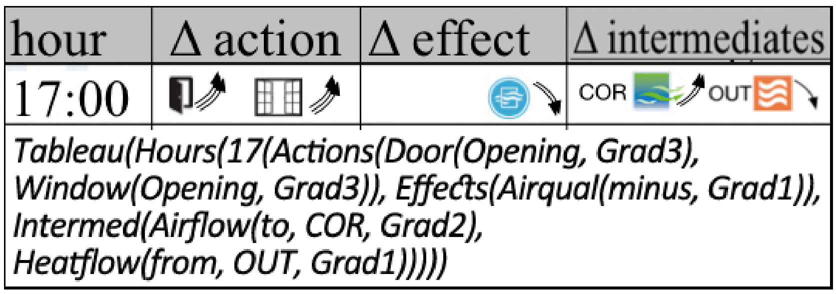


Figure 6.1: Table and its linearized expression

- MA, Morphological Analysis, implements a string-to-tree transducer in order to produce a decorated tree, the only data structure that can be manipulated by the Ariane-Heloise system. It also distinguishes between strings corresponding to "concept-words" of the table and the elements of the tree structure (parentheses and commas).

[1] jean-philippe.guilbaud@imag.fr

- XA is an analysis dictionary that contains all the concepts used in the tables with their description by means of variables values.

- SA1, Structural Analysis 1, produces the exact image of the table and its contents as a decorated tree.

- SA2, Structural Analysis 2, modifies the tree produced by SA1 to give it a syntagmatic linguistic expression, ready to be translated into French.

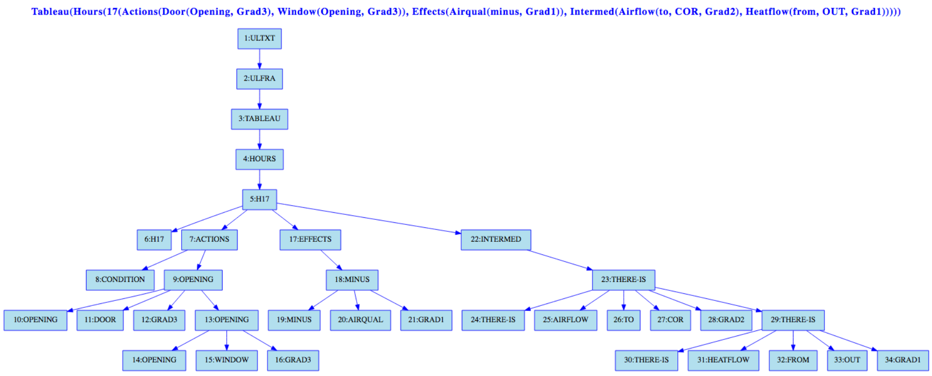


Figure 6.2: SA2

- LT, Lexical Transfer, translates the words of the table into French and sets the syntactic functions of words and phrases.

- SG, Syntactic Generation, calculates the surface syntactic structure of the sentences of the French target text and the morphological agreements in person, gender and number between the lexical constituents of the linguistic tree.

- MG, Morphological Generation, implements a tree-to-string transducer which first produces the correct forms of the words of the leaves of the linguistic tree and then outputs the final French text corresponding to the table contents.

Traduction

Dans le créneau horaire 17h-18h, si vous aviez laissé la porte et la fenêtre ouvertes beaucoup plus longtemps, la qualité de l'air aurait diminué un petit peu, il y aurait eu un courant d'air sensible vers le couloir et un flux thermique léger venant de l'extérieur ;

Figure 6.3: Output from Ariane-Heloise GRA-FRA GM phase

6.2 Human-machine interface

Given the explanation engine and the text generation, a user interface (UI) is mandatory to convey the explanations and to involve the inhabitant in the energy management process. However, displaying information and providing an interactive system is not enough to involve the inhabitant in such a process. If the user is unaware, or not motivated, or unable to control his energy consumption, such a system is useless. Therefore, a UI has been designed for the e-consultant system that supports awareness, aimed at persuading and helping inhabitants to change their behavior in order to reduce their energy consumption while feeling comfortable. The following focuses on such a UI and illustrates the design principles. The design and implementation of the system was done in cooperation with the IIHM (Ingénierie de l'Interaction Homme-Machine) team of the LIG laboratory and is part of VAN BAO NGUYEN's PhD[2]

6.2.1 User interface requirements

With the industrial partners (Elithis, and Vesta Systems), the e-cosultant is intended to be deployed on a tablet installed on a wall of the living room or in the entrance hall of each apartment in a tower being built at Strasbourg, France, Figure 6.4, composed of 63 apartments and 800 square meters of offices and commercial space. Thus, several issues have to be considered: (i)

[2] van-bao.nguyen@univ-grenoble-alpes.fr

standing in front of the tablet to use it may become quickly uncomfortable compared to sitting at a desk; (ii) depending on the time of day, the user may not be available to interact with the system (e.g. leaving home in the morning to go to work); (iii) over time, users may not pay attention to the device any more.

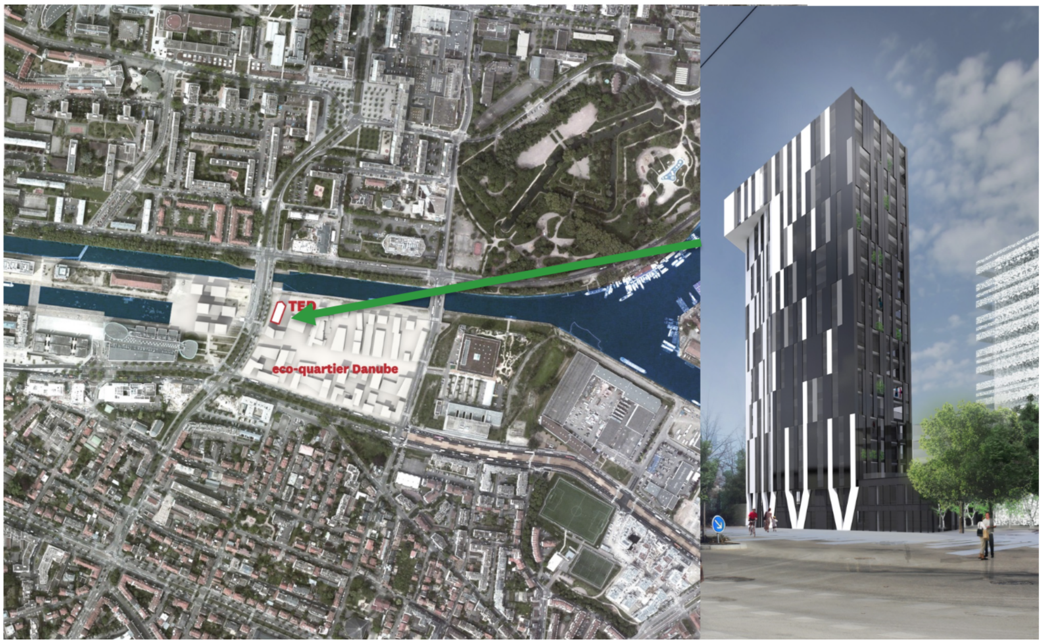


Figure 6.4: Elithis tower in Strasbourg, France

To address these issues, the UI must satisfy the following two design requirements: The UI must be eye catching, Multiple levels of interaction for different contexts of use (e.g. key moments of the day) must be supported. These two requirements aim at maintaining high motivation and ability. For example, by embedding utility functions such as time and weather forecast, the UI should catch the eye at key moments of the day (e.g. looking at the clock before leaving home). Consequently, it should maximize (i.e. make the user aware of) the persuasive features of the system, such as reminders of user chosen targets, as well as greetings and rewards in order to maintain appropriate behaviors.

6.2.2 Designed for daily life: catching the eye at the right time

One compelling approach to potentially impact the motivation of individuals is to offer an interactive system for daily life. Our suggestion is to bring daily activities and objects, such as consulting the time, checking the weather or decorative elements, into the UI. Figure 6.5 illustrates the design concept in a "Mondrian" style. To satisfy requirement, it is hypothesized that an artistic-like UI (e.g. a painting) embedding daily utility services (e.g. a clock) would catch the eye at any moment of the day and be attractive.

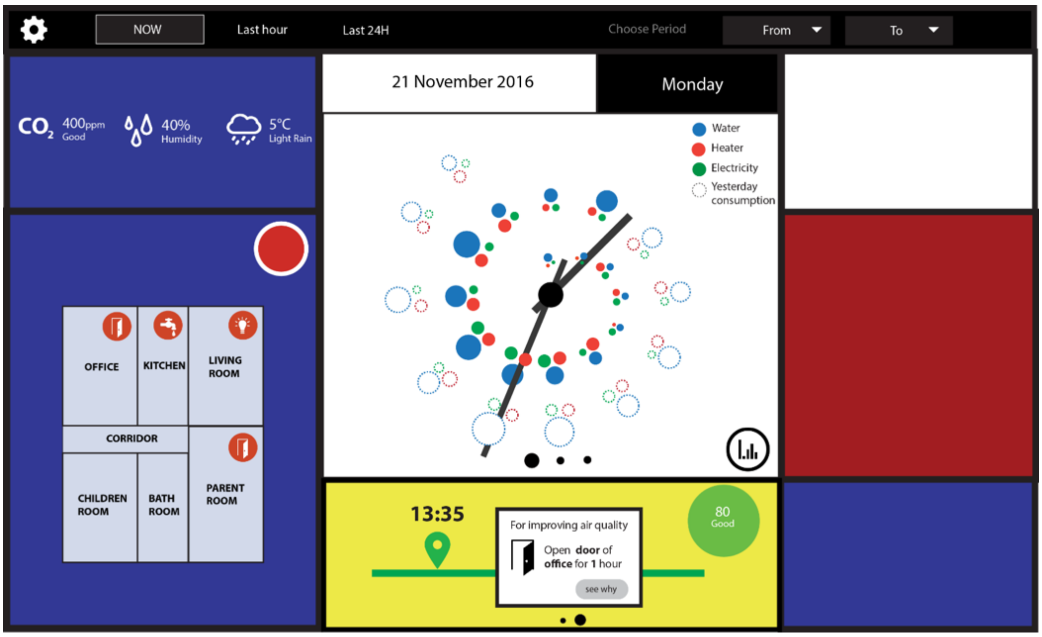


Figure 6.5: Design concept inspired by Mondrian's style

6.2.3 User interface in detail

The tiles, in figure 6.5, represent a block of information that corresponds to a specific activity. At the center of the interface, a 24-hour clock provides an overview of the energy consumption of the last 24 hours. At a glance, a zoomed-out tile provides synthesized information that requires no or very little user interaction. If more detail is needed, using semantic zoom interaction techniques, a tile can be expanded providing more space at the expense of the other tiles. Three major tiles are identified: a spatial view of the habitat (blue tile on the left hand side), a temporal view of the energy consumption (clock-based white tile), and the e-consultant view (yellow tile). The white and yellow tiles both support sliding gestures to present complementary views. On the left hand side, the blocks filled in blue provide information about the outside conditions (top-left blue rectangle: air quality, humidity percentage, weather conditions) and about the inside of the habitat. The latter is represented through a map of the home where a rectangle represents a room or a corridor. As detailed in the following, red and green circles indicate the overall status of the home as well as the status for each room. At the center of the screen, the white tile displays a clock (a utility service) associated with time-based synthetic visualizations. The spiral-based visualization presented here is only one possible illustration. Additional complementary visualizations can be obtained by a sliding gesture performed at the bottom of the tile. These visualizations represent information about energy consumption such as domestic activities, or daily objectives. The yellow tile (center-bottom) is dedicated to the interaction with the e-consultant. For instance, a synthetic view may be a representation of the suggested next actions to be achieved within the current hour slot of the day. With a click, the user can obtain detailed explanations about the action and its consequences.

Another feature of the UI is the use of everyday life metaphors. When people face something new and want to understand it, they usually try to

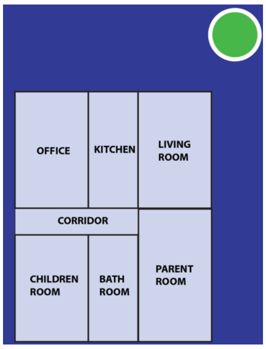


Figure 6.6: No issues: green light

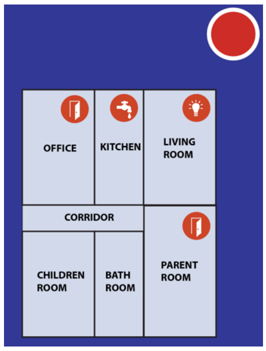


Figure 6.7: issues highlighted with a red icon

compare it to things that they already know in order to fit it into their knowledge structure. Thus, the use of everyday life metaphors in a UI may facilitate learning. The metaphors can be based on an activity or an object that is familiar to the user. The UI can have used the traffic light metaphor to promote a glanceable (i.e. concise) UI: "at a glance, I can check if everything is OK in my home". As explained above, this part of the UI (Figure 6.5) represents a logical representation of the habitat (less accurate than a real 2-dimensional map). The traffic light metaphor is used as follows: if the system detects an undesired event (i.e. lights are on in a room with no-one present), a red circle (Figure 6.7) is drawn in the related room associated to an icon (e.g. a light bulb means that lights are on; an open door means that, according to the e-consultant, a door should stay closed in order to optimize temperature in the room and/or to maintain air quality). The circle located at the top-right corner provides an overview. Therefore, a green circle means that everything is OK (Figure 6.6).

The UI is designed for everyday tasks. Another method to keep users motivated is to give them full and easy control of their habitat. It is aimed at simplifying the complex notion of multi-criteria optimization. The solution is to transform these tasks into the form of an everyday activity such as navigating on a map or manipulating familiar widgets. Figure 6.8 shows how the use of familiar widgets such as sliders, allows users to find the preferred trade-off between thermal comfort, cost and energy waste.

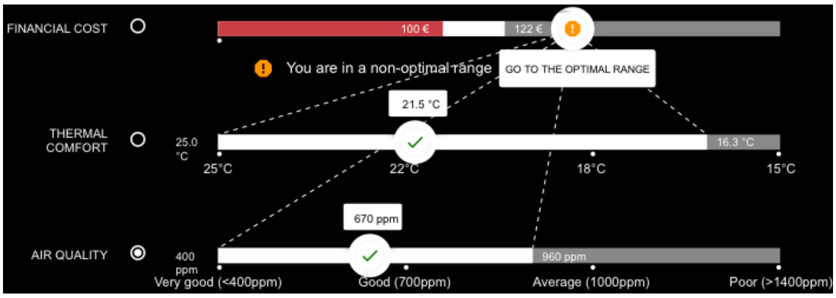


Figure 6.8: Finding the appropriate compromise between conflicting criteria using the Trade-Off-Pareto sliders

The trade-off between the criteria once defined, the e-consultant engine provides a list of recommended actions. Figure 6.9 shows the actions plan that is recommended for the day. It represents the optimal scenario computed by the explanation engine presented earlier in the fourth chapter. Clicking on the "see why" button attached to a particular action gives access to explanations about the action and its consequences, figure 6.10.

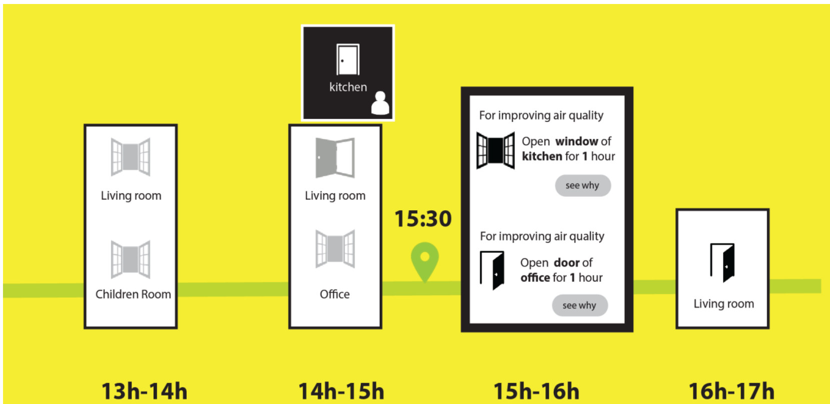


Figure 6.9: The actions plan recommended for the day

This way the user can get access to the explanations in ease and obtain the explanations for the action he interrogate. The explanations will appear in the natural language easy to read and understand by occupants.

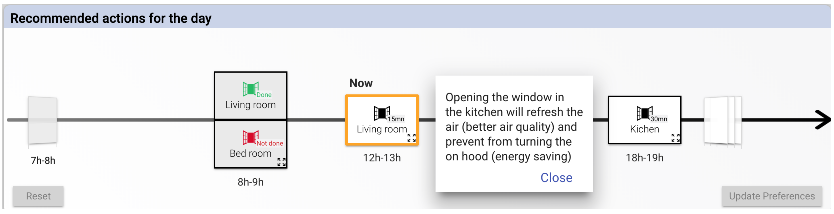


Figure 6.10: Recommended action with explanations

This section has described the interactions between the user and the energy system. It also described the main features that need to be present in the UI to help the user to control his energy consumption sensibly.

Next section describes the field studies, that had been done to validate the explanations and their utility.

6.3 Validation scenario for the generated explanations

6.3.1 Context and goals

Introduction to the validation scenario:

" Homes are complex systems where different phenomena from nature are present. Occupants generally would like their comfort criteria (like thermal comfort and air quality) to be at the optimum without increasing their consumption or energy bill. Yet, within these complex systems, occupants have difficulty in determining the optimal set of actions they should perform within a specified context or being able to estimate the impact of their actions. For this reason this work proposes an assisting tool ("the explanation generator") to help occupants to understand the impact of their actions and the cause-effect relations between different variables within different contexts."

The objective is to measure how the causal explanations proposed in this work might assist occupants to better understand their homes. For that, three criteria need to be satisfied, i.e. are the generated explanations:

1. Intelligible
2. Credible
3. Easy for the user to understand them

6.3.2 Method

The method aims to evaluate if the proposed tool to generate explanations can help the users to understand their energy systems or not. To do that the first step will immerse the participant in the scenario: an office (the case study) with an energy system to evaluate their initial knowledge of the system and to aid them in thinking deeper into the problem. This is measured through the first task (a practice for the occupants to do). The explanations are given to participants then they are given the second task to evaluate the utility of the explanations. Finally, a registered semi-structured interview takes place with the participant to get his feed back about:

- If the participant finds the explanations intuitive or not.
- If the explanations are clear or not.
- If the participant is ready to adopt them or not.
- The form of explanations.
- Other comments.

Their feedback is analyzed to determine their understanding of the different phenomena.

6.3.3 Participants

The 10 participants were between 18 and 65 years old, 5 women and 5 men. The participants were from different backgrounds (scientific and non-scientific), none of them from the domain of the research and never worked on the problem of energy management. They were also volunteers and were not paid for their participation.

6.3.4 Independent variables

In order to get a valid comparison all the tests were done in an office in the G-SCOP laboratory, and with one day from the history, 05/05/2015 (chosen randomly), as different days may be clearer or more complicated for participants so that their answers won't be comparable. All the participants were asked to perform the same actions: opening of the window and door. The measured variables were the proposed programs for the opening of door/window throughout the day. The results were evaluated using a physical model and evaluated by how much they improved the comfort of occupants.

The participants were interviewed by a researcher from the human and social sciences (Hélène Haller [3]) not directly involved in this research. All interviews were recorded via two microphones and then analyzed.

6.3.5 Tasks

The participants were asked to perform two tasks. In Task 1 (T1), the participants were instructed to look at data from the 5^{th} of May 2015 with

[3] helene.haller@umrpacte.fr

the context variables for the case study. Then they were asked to determine the best window/door position to enhance their comfort. Task 2 (T2) was repeating the first task after having been given the system's explanations. Then a short semi-structured interview with each participant was done to see if they had any preference or any comments on the explanations.

6.3.6 Scenario of the interview

This section describes the exact speech and questions presented to the different participants:

We thank you for agreeing to participate in this experiment. It is part of a doctoral work on the realization of an energy management system. This system is based on the generation of explanations, containing tips for users. Today, we wish to observe whether the proposed explanations are understandable and acceptable for you.

For this, we offer a scenario, during which we will ask you several questions. (Give figure 6.11 - office and sensors) The questions we will ask you are related to this office, located in the laboratory G-SCOP. This office is equipped with different sensors. The only possible actions to act on this room are to open or close the window and the door. Now imagine that we are 5 May 2015. (Give image 6.12 - Office registered data for the 5 May 2015) Here are the data provided by the sensors installed in the office, data relating to the air quality, the number of occupants, the outside temperature, the corridor temperature and the office temperature at 8 am.

Given these data, what do you think are the best action (Opening door /window) to get the maximum comfort in this office? Generally, comfort is defined by a temperature between 21^o and 23^o C and a CO_2 concentration as low as possible. I will now show you how schematically how the generation of direct explanations works. (Give picture 6.13 - model for direct explanations)

Figure 6.11: office and sensors

I will let you read it.

Given this pattern, what would you do to get the maximum comfort in this office? I remind you that comfort is generally defined by a temperature between 21^o and 23^o C and a CO_2 concentration as low as possible.

- Do you find that the direct explanations I showed you were logical or not? Why ?
- Have these explanations helped you better understand how to get maximum comfort in the office? Why ?
- Do you find this type of explanation (direct explanations) intuitive / understandable for you?
- Do you think that these explanations could be better presented? If yes,

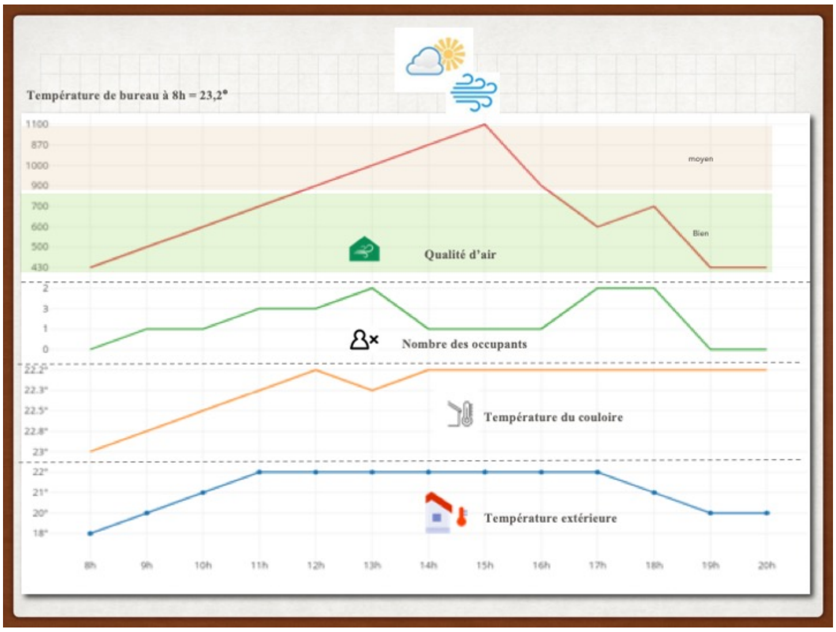


Figure 6.12: Office registered data for the 5 May 2015

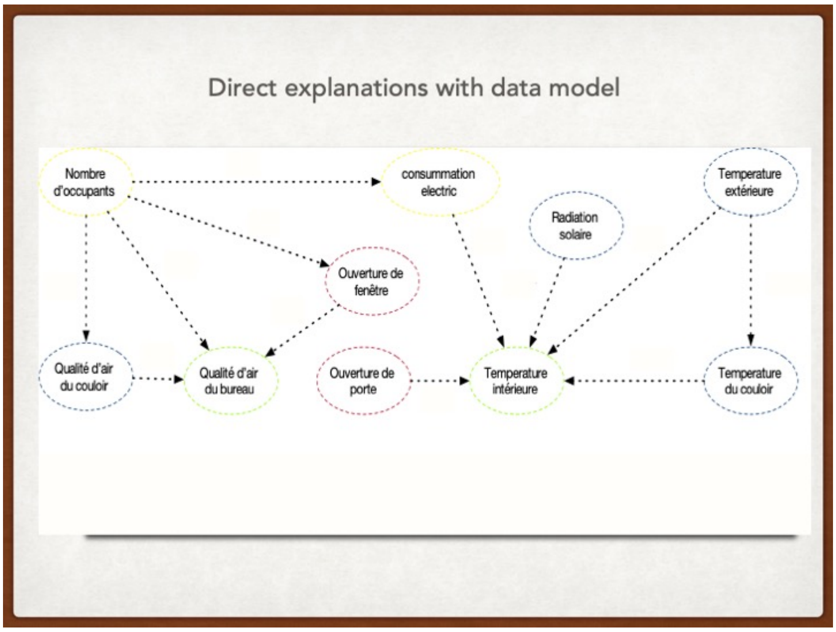


Figure 6.13: Direct explanations

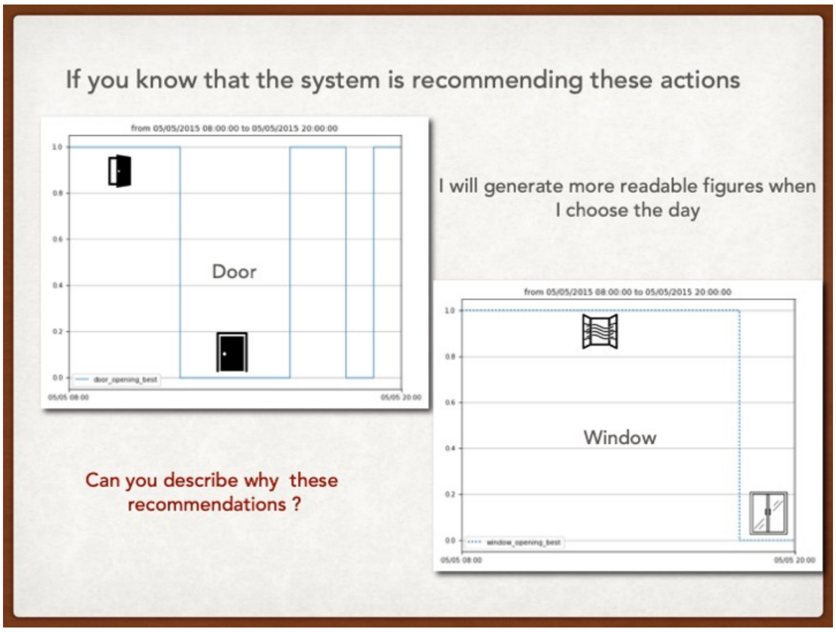


Figure 6.14: Recommended solution

how ?

(Give picture 6.14 - Recommended solution) Here are the actions recommended by the system for maximum comfort. Do you understand these recommendations? I will now show you a differentiated causal explanation (Give picture 6.15 - Differential explanations). Between 8 am and 9 am, if you left the window and the door open longer, there would have been a light heat input from the outside and from the corridor, which will improve your thermal comfort in the 9h-10h time slot and improve the air quality in the time slot 10h-11h.

- Do you find these explanations logical ? Why ?
- Have these explanations helped you better understand how to get maximum comfort in the office? Why ?
- Do you find this type of explanation (causal explanations) intuitive?

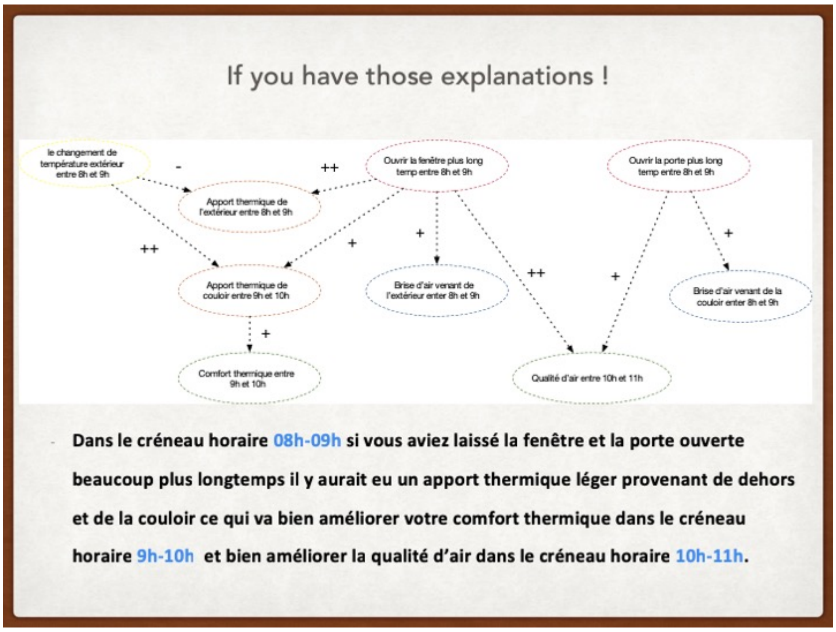


Figure 6.15: Differential explanations

- Do you think that these explanations could be better presented?
- Have explanations and understandings encouraged you to follow the recommendations of the system?
- Have these explanations increased confidence in the system?
- Could you give an estimate of the improvement of your understanding of the environment following these explanations?

Before concluding this interview, I would like to ask you two questions:

- By comparing the two types of explanation (direct explanation / differentiated causal explanation), which one do you find best? Why ?
- If you had these explanations at home, would you use them? If so, how often ? for what occasion ?

6.3.7 Results

In the first task (T1), participants in general were a little bit lost and tried to imagine scenarios that were far from what the system would recommend them to do (the best solution). In general, they based their decisions on their habitual activities or certain beliefs (like "we should open the window each morning"), except for one participant who tried to analyze the context data and seemed to have a better understanding. After showing the direct explanations, participants were asked to complete the task (T2). A clear improvement was noticed, as they (from what they reported) started to know what the variables affecting the comfort variables (inside temperature, air quality) were. So they had started to realize the relation between actions and resulting effects.

Then, the differential explanations were introduced with the optimal solutions. Finally, the participants were asked to compare the differential and direct explanations and give their opinions.

Participants in general did appreciate the explanations and most of them repeated the same words "I learned new things" or "I didn't know that before" or "It confirmed what I thought". Around 10% of participants preferred the direct explanations, others preferred the differential explanations and also liked the natural language form of the explanations. In remarks, they said in general that "heat flow" is not clear for them and asked for an easier term to replace it.

Participants said that they would like to get the explanations and system recommendations, even if they would not follow them all the time. They will consult the explanations more when they suspect that something is not right (like to checking it when they feel cold even when the heater is on), or, in the case of pollution periods, know what to do. One participant said that she will check it each day to confirm the air quality for her children.

6.4 Second Validation

A second field study by the IIHM team at the LIG laboratory was done to evaluate the UI and the causal explanations that were part of it after being integrated.

This study was done with a different interviewer, different building and with different participants from the previous validation study. 13 participants from different backgrounds, age and sex were chosen. For the explanation part, all of the participants well understood the explanations provided by the e-consultant and most of them found the explanations well formulated; results are presented in the table:

	Yes		NO	
	Value	Percentage %	Value	Percentage %
Could you understand the explanations provided by the system?	13	100	0	0
Would you formulate the explanations differently?	2	15.3	11	84.7
Do you find the explanations useful?	13	100	0	0
Do you think that explanations are necessary to understand how the e-consultant works?	8	61.5	5	38.5
Would you find it useful to provide the system with explanations regarding your behavior?	10	76.9	3	23.1

Only two participants declared they would formulate explanations differently. Nevertheless, they used different terms to express their understanding of the purpose of explanations such as: 'reasons' (Participant 1), 'explain(ations)' (Participants 2, 5, 7, 11, 12 and 14), the 'Why' (Participant3), 'consequences' (Participant 4 and 14), 'motivator' (Participant 6), 'utility' (Participant 8). All participants declared they found the explanations useful and a majority (8/13) found them necessary in order to understand how the e-consultant works. In addition to usefulness, explanations appeared to contribute generally in a positive way regarding the differences as highlighted by the verbatim report below (some of them are in French and translated into English):

1. Answer:
   « C'est exactement ça, on te demande de faire des trucs. Du coup, par défaut, ça m'énerve parce que je n'ai pas envie de faire des trucs ; et, du coup, mais il m'explique ; je vais t'expliquer pourquoi quoi. Donc, ça, c'est bien. Du coup, quand il m'explique comme ça ; moi, après, je comprends et je dis OK . . . Et ben, c'est directement lié aux motivations quoi. Donc, si je suis motivé par la raison, si je n'en ai rien à foutre . . . C'est pour ça que j'imaginerais la possibilité de lui dire mes motivations. S'il sait exactement la température que je préfère, ça, c'est parfait. Du coup, je n'ai pas trop à m'en occuper ». (Participant 2)
   English translation:
   "That's exactly it, you're asked to do things. So, by default, it annoys me because I do not want to do things; and suddenly, but he explains to me; It explain why. So, that's good. So, when it explains me like that; me, after, I understand and I say OK ... Well, it's directly related to what motivations. So, if I am motivated by reason, if I do not give a fuck ... That's why I would imagine the possibility of telling him my motives. If he knows exactly the temperature that I prefer, that's perfect. So, I do not have too much to take care of it ». (Participant 2)

2. Answer:
   « Oui, parce que la première fois, tu as envie de savoir pourquoi. Est-ce que c'est par rapport à ce que tu penses ; est-ce que ça confirme tes attentes ? Des fois, c'est une autre raison. Voilà, c'est pour conforter l'utilisateur ». (Participant 5)
   English translation:
   «Yes, because the first time, you want to know why. Is it in relation to what you think; does that confirm your expectations? Sometimes, that's another reason. That's it to comfort the user »(Participant 5)

3. « This is, I guess, some information that makes you motivated. Why should I do this now? You will have your answer ». (Participant 6)

4. Answer:
   « En fait, ça te donne une explication . . . L'explication qu'il y a derrière l'action, derrière son conseil . . . parce qu'en fait, te dire : ferme la fenêtre, ouvre la fenêtre, mais si tu ne sais pas pourquoi tu le fais, ça peut te . . . A un moment, tu peux te dire : pourquoi je le fais ; tu peux t'arrêter mais, quand tu vois une explication, en plus qui est plausible, qui tient la route, tu vas te dire : je le fais ». (Participant 11)
   English translation:
   «In fact, it gives you an explanation ... The explanation behind the action, behind his advice ... because in fact, tell you: close the window, open the window, but if you do not know why you can do it, it can ... At a certain moment, you can say to yourself: why I do it; you can stop but, when you see an explanation, besides which is plausible, who holds the road, you will say to you: I do it ». (Participant 11)

These two field studies demonstrate the importance of explanations for occupants. They present the explanations' utility in allowing the occupants to understand how the environment is functioning and why the e-consultant is recommending different actions at different times.

6.5 Conclusion

This chapter has presented how the explanations of the EMS are integrated with the e-consultant. This was done with the aim of creating a channel of cooperation between occupants and the EMS through the explanations.

It also explains how to transform the causal explanations into a natural language, which is done to facilitate the transfer of explanations to occupants.

The second part presented the human-machine interface, which is the mechanism of interacting with the occupants and presenting the explanations to them.

Finally, it presents the field studies to validate the direct/differential explanations' usefulness and acceptance by occupants.

Chapter 7

General conclusion

Energy management systems (EMSs) were designed and constructed in the beginning with the aim of fully automating and optimizing energy consumption. One primary driver behind EMSs is to increase the occupants' satisfaction within their environment, without extra cost, and reduce the waste of energy. Human activities cannot be neglected and are part of the system state. For example, an occupant may adjust the thermostat, open the window, or turn on the lights, to achieve objectives that might be hidden or ambiguous to the EMS. Therefore, EMSs have to be highly cooperative human-machine systems. Indeed, the EMS provides services to occupants but occupants are also part of the system and influence it significantly with their own behavior. Therefore, advanced EMSs need to cooperate with the user to help him to achieve his objectives. Additionally, innovative end-user services, such as replaying past situations, anticipating the future, or mirroring the current state, are also important. With the goal of enhancing the cooperation between humans and EMSs, this work focuses on explanations as an important tool to achieve this cooperation.

The beginning of the work starts by explaining why energy consumption in buildings is economically crucial and a key element to optimize the reduction in energy waste. Then it presents the different levels of knowledge between the occupants and energy systems. It also describes why it is hard for occupants to easily understand the different physical phenomena present in the environment, and at the same time, why it is hard to understand how the EMS is working and making decisions.

The work continues by exploring the different energy systems and compares them, presenting the difficulties when trying to generate causal explanations to convey the knowledge present in the EMS to occupants.

The scope of work required a search for the different systems and methods usually used to generate explanations, like expert systems, then showing what the useful parts are that can be reused or adapted with the EMS and why.

Next, this thesis proposes a new method to generate differential explanations based on the comparison of different scenarios. By means of a real case study, it shows how the differential explanations are calculated and how the causal graph is obtained.

It continues by creating a new method to generate explanations when the model-based EMS cannot be deployed and used. It discusses the difficulties in using differential explanations without an energy model, and then proposes a method to build and validate a data model. The objective is to be able to compare different days based only on their measured variables. Next, it describes the different steps needed to achieve this concept and the limitations encountered when applying differential explanations with it. Therefore, it proposes a new way of generating explanations called, direct explanations, by learning a Bayesian network from similar days.

The last chapter presents a validation method with a field study to confirm the result different studies on the utility of causal explanations and their adoption by occupants with the EMS. Then it gives a brief description of how the causal explanations are transformed to text in natural language and why this step is useful for occupants. Finally, it shows how the explanations are going to be integrated into a human-machine interface to get the best use out of them and ensure a better and more agreeable experience for the occupants.

From the technical perspective, an early prototype of the INVOLVED e-coach system has been developed that provides inhabitants with a 24-hour plan of recommended actions along with contextual differential explanations that justify each action. This plan satisfies the user's preferred compromise between thermal comfort, air quality, and financial cost specified by the user. Users can edit the plan, e.g. suppress an action, skip some actions, perform additional actions, or even change their preferred compromise, and be informed in real time of the consequences on energy consumption and comfort. In practice, in the prototype's current version, the generation of explanations has been tested while deployed in a controlled environment (the case study), and the user-centered evaluation of the user interface elements has also been performed in a controlled setting where the behavior of the explanations generator was simulated. Whereas these early steps are necessary to detect the basic technical flaws and limitations of a system, they must be completed by longitudinal experiments performed in real world settings. This part is under discussion with the industrial partner in the Elithis Tower and in some other buildings. Based on previous research results and field studies on home automation, it is expected that users will be interested in the system recommendations in the very first months of use, provided that they are useful, robust, and trustworthy. In the long run, either the e-coach will not be used anymore, or the settings (compromises) will be good enough and the system will run in the background as calm and optimal technology until some exception occurs. In this case, the INVOLVED system could be called upon to specify a new compromise relevant to the current situation, but there is no guarantee that the user will be motivated enough to address the exception. Clearly, these hypotheses on future use will need to be validated experimentally.

Chapter 8

List of publications

Journals

1. Unmasking the causal relationships latent in the interplay between occupant's actions and indoorambience: a building energymanagement outlook
   Monalisa Pal, Amr ALZOUHRI ALYAFI, Stéphane PLOIX, Patrick REIGNIER, Sanghamitra Bandyopadhyay
   Applied Energy volume 238, 15 March 2019, Pages 1452-1470

2. From Usable to Incentive-Building Energy Management Systems
   Amr Alzhouri Alyafi, Van Bao Nguyen, Yann Laurillau, Patrick Reignier, Stéphane Ploix, Gaëlle Calvary, Joëlle Coutaz, Monalisa Pal, Jean Philippe Guilbaud
   Modeling and using context journal 2018, Open science

International Conferences

1. Enhancing Comfort of Occupants in Energy Buildings
   Monalisa Pal, Amr Alzouhri Alyafi, Sanghamitra Bandyopadhyay, Stéphane Ploix, Patrick Reignier
   Operations Research and Optimization, pp.1-11, 2018.

2. Analysis of Optimizers to Regulate Occupant's Actions for Building Energy Management.
   Monalisa Pal, Raunak Sengupta, Sanghamitra Bandyopadhyay, Amr Alzouhri Alyafi, Stéphane Ploix, Patrick Reignier, Sriparna Saha
   ICAPR 2017 - Ninth International Conference on Advances in Pattern Recognition,
   Dec 2017, Bangalore, India. IEEE, IEEEXpolre.

3. Explanations Engine For Energy Management Systems in Buildings
   Amr Alzouhri Alyafi, Jean-Phillipe Guillbaud, Patrick Reignier, Stephane Ploix
   The 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications,
   Sep 2017, Bucharest, Romania. IEEEXplore

4. Differential Explanations for Energy Management in Buildings
   Amr Alzouhri Alyafi, Monalisa Pal, Stephane Ploix, Patrick Reignier, Sanghamitra Bandyopadhyay
   Computing Conference 2017, Jul 2017, London, United Kingdom. IEEE, IEEEXplore.

Chapter 9

Appendix

9.1 Optimization criteria

The optimization problem aims to minimize various factors contributing to the occupant dissatisfaction ($\mathcal{S}_{avg}$ of equation (9.1)) when there is at least one occupant present in the room (i.e. $n^k \neq 0$). The occupancy factor is considered so as to avoid generating recommendations when there is no one present in the room like on holidays. The proposed work characterizes the effect of occupant actions with respect to the following criteria:

1. thermal dissatisfaction (σ_{temp}^k) of equation (9.2) where T_{in}^k is obtained in Kelvin[1] (or K) and n^k is the average estimated occupancy at the k^{th} hour,

2. dissatisfaction associated with CO_2 based air quality (σ_{air}^k) of equation (9.3) where C_{in}^k is obtained in μmol per mol of air[2] and n^k is the average estimated occupancy at the k^{th} hour,

3. an indicator associated with scaled expenditure incurred by energy

[1] 273.15K = 0 degree Celsius

[2] 1 μmol per mol of air = 1 parts-per-million particles of air

consumption inside the office room i.e. σ^k_{cost} of equation (9.4) where P^k_{elec} is obtained from $\mathcal{H}$, P^k_{fuel} is obtained and n^k is the average estimated occupancy at the k^{th} hour, and

4. the annoyance caused by recommendations for hourly change in window and/or door status (δ^k_{WD} of equation (9.5) which is a unit-less quantity).

$$\begin{aligned}\mathcal{S}_{avg} &= \frac{1}{24}\mathcal{S} = \frac{1}{24}\left[\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \mathcal{S}_4\right] \\ &= \frac{1}{24}\left[\sum_{k=0}^{23}\sigma^k_{temp}, \sum_{k=0}^{23}\sigma^k_{air}, \sum_{k=0}^{23}\sigma^k_{cost}, \sum_{k=1}^{23}\delta^k_{WD}\right]\end{aligned} \tag{9.1}$$

where, at every k^{th} hour,

$$\sigma^k_{temp}\left(T^k_{in}\right) = \begin{cases}\frac{294.15 - T^k_{in}}{294.15 - 291.15}, & \text{if } T^k_{in} < 294.15 \text{ and } n^k > 0 \\ 0, & \text{if } 294.15 \le T^k_{in} \le 296.15 \text{ or} \\ & n^k = 0 \\ \frac{T^k_{in} - 296.15}{299.15 - 296.15}, & \text{if } T^k_{in} > 296.15 \text{ and } n^k > 0\end{cases} \tag{9.2}$$

$$\sigma^k_{air}\left(C^k_{in}\right) = \begin{cases}0, & \text{if } C^k_{in} \le 400 \text{ or } n^k = 0 \\ \frac{C^k_{in} - 400}{1500 - 400}, & \text{if } C^k_{in} > 400 \text{ and } n^k > 0\end{cases} \tag{9.3}$$

$$\sigma^k_{cost}\left(P^k_{elec}, P^k_{fuel}\right) = \begin{cases}\frac{P^k_{elec}E_{elec} + P^k_{fuel}E_{fuel}}{1000}, & \text{if } n^k > 0 \\ 0, & \text{if } n^k = 0\end{cases} \tag{9.4}$$

$$\begin{gathered}\delta^k_{WD}\left(\zeta^k_{pair}\right) = \begin{cases}\delta^{k-1}_{WD}\left(\zeta^{k-1}_{pair}\right) + 1, & \text{if } \zeta^k_{pair} \ne \zeta^{k-1}_{pair} \\ \delta^{k-1}_{WD}\left(\zeta^{k-1}_{pair}\right) + 0, & \text{if } \zeta^k_{pair} = \zeta^{k-1}_{pair}\end{cases} \\ \text{with } \zeta^k_{pair} = \left(\zeta^k_W, \zeta^k_D\right) \text{ and } \delta^0_{WD}\left(\zeta^0_{pair}\right) = 0\end{gathered} \tag{9.5}$$

In many cases, the concerned objectives are conflicting in nature. The following points provide some insights into the degree of conflict between the objectives:

1. *Between σ_{temp} and σ_{air}*: On a winter day, when T_{out} is lower than desired T_{in}, opening of windows can lead to lowering T_{in} and in turn, increasing σ_{temp}. However, the same action might also be accompanied with decreasing C_{in} and in turn, decreasing σ_{air}. Thus, an occupant is baffled with the choice of actions. It should be noted that although the underlying physical phenomena are not such simple, deterministic if-then rules, yet the example scenario is efficient at describing the conflicting nature of the objectives.

2. *Between σ_{temp} and σ_{cost}*: On a winter day, when T_{out} is extremely lower than desired T_{in}, closing of windows can lead to increasing T_{in} and in turn, decreasing σ_{temp}. However, due to thermal inertia, the desired σ_{temp} cannot be attained instantly. In such a situation, occupants will be tempted to turn on the heater [69] and in turn, will incur associated expenses due to energy consumption which will increase σ_{cost}.

3. *Between σ_{temp} and δ_{WD}*: As an example, in order to regulate T_{in} and hence, decrease σ_{temp}, let an optimal schedule correspond to several changes in door and window status like around 10 toggles between open and close. However, occupants can forgo a certain amount of σ_{temp}, instead of getting recommendations throughout the day which gets in the way of their daily routine. Hence, decreasing δ_{WD} results in an increase of σ_{temp} and a correct balance is very much essential.

4. *Between σ_{air} and σ_{cost}*: Sometimes closing windows have a detrimental effect on σ_{air} and hence, in order to lower σ_{air}, an alternative occupant action is opening the door which facilitates air flow between the room

and the neighboring zones. However, in such a situation, the targeted volume for the heater drastically increases which leads to keeping the heater on for a longer duration and thus, increases σ_{cost}.

5. *Between* σ_{air} *and* δ_{WD}: Similar arguments used to demonstrate the conflicting nature of σ_{temp} and δ_{WD} can also be used to establish the conflicting nature of σ_{air} and δ_{WD}.

6. *Between* σ_{cost} *and* δ_{WD}: As an example, in order to regulate T_{in} and C_{in}, an optimal schedule can correspond to several changes in door and window status. However, to reduce interference with occupants' daily routine, the proposed approach tries to minimize δ_{WD}. But this prevents T_{in} to reach the optimum and/or the desired value. When it is necessary to counteract this effect, the heater can be used to balance the T_{in}, which, in turn, increases the σ_{cost}.

The above-mentioned examples demonstrate the conflict among the objectives used in the proposed work and such conflict among the objectives necessitates the use of a multi-objective optimization algorithm to address this problem. As there are more than three objectives, this optimization problem belongs to the special class the many-objective optimization problems [54].

9.1.1 Compromise of interest

On solving the unconstrained many-objective minimization problem with multiple conflicting and incommensurable objectives, a set of solutions ($\mathcal{A}^{PS}$) is obtained as the estimated Pareto-optimal solutions [54]. These solutions represent different degrees of trade-offs among the objectives. The set of objectives corresponding to the estimated Pareto-optimal solutions form the estimated Pareto-front ($\mathcal{S}^{PF}_{avg}$) [54]. A single solution is considered from this set of Pareto-optimal solutions either by an automated selection or by a user intervention.

Automated decision making (the default paradigm)

Assuming the concerned problem requires a solution that has a preference weighting of w_1, w_2, w_3 and w_4 on the four objectives of equation (9.1), the decision-making is governed by equation (9.6) which is the solution with the minimum weighted city-block distance [80] of solutions constituting the estimated Pareto-front from a reference objective vector ($\mathcal{S}_{avg}^{ref}$). For a minimization problem, where the range of objectives is the first hyper-octant of the real-valued space, the reference objective vector for decision making corresponds to the origin i.e. 0 for each objective. Thus, from a mathematical standpoint, the final optimal objective ($\mathcal{S}_{avg}^{\star}$) corresponds to the discovered point closest to the origin of the objective space. On the other hand, from the perspective of building systems, the final optimal objective ($\mathcal{S}_{avg}^{\star}$) corresponds to the point with the minimum net (global) occupant dissatisfaction. To illustrate this strategy, the solutions constituting the Pareto-front out of several possible states are shown in figure **??**, along with the selected Pareto-optimal solution ($\mathcal{S}_{avg}^{\star}$) closest to reference objective vector ($\mathcal{S}_{avg}^{ref}$) in terms of city-block distance.

$$
\begin{aligned}
\mathcal{S}_{avg}^{\star} &= \underset{\mathcal{S}^{PF}}{\arg\min} \sum_{i=1}^{4} \left(w_i \times \left| \mathcal{S}_{avg,i}^{PF} - \mathcal{S}_{avg,i}^{ref} \right| \right) \\
&= \underset{\mathcal{S}^{PF}}{\arg\min} \sum_{i=1}^{4} \left(w_i \times \mathcal{S}_{avg,i}^{PF} \right) \text{ with } \mathcal{S}_{avg}^{ref} : (0,0,0,0) \\
&\text{where, } \sum_{i=1}^{4} w_i = 1 \text{ and } \mathcal{A}^{\star} = \arg\left(\mathcal{S}_{avg}^{\star} \right)
\end{aligned}
\tag{9.6}
$$

For this work, the preference weights of the objectives are defined as $w_1 = 33.22\%$, $w_2 = 33.22\%$, $w_3 = 33.22\%$ and $w_4 = 0.34\%$. This implies that thermal dissatisfaction, air quality dissatisfaction, and energy cost indicator are targeted equally to be minimized whereas, the importance of minimizing changes in actions is very small. Restricting changes in actions will prohibit the other optimization objectives to evolve. So, until several other kinds of occupants' actions are integrated with this framework, in future, the fourth objective has the least importance from a decision-making perspective.

It should be noted that if in equation (9.6), $\mathcal{S}_{avg}^{ref}$ is set to the objective values for the actual scenario, the system can generate a solution that causes the minimum overall change in satisfaction, yet is a Pareto-optimal solution. Thus, if the occupant highly prefers the objective or satisfaction values according to the actual scenario, this strategy of decision making can also be used to find a Pareto-optimal solution with the minimum overall change in satisfaction. Nonetheless, in this work, as the default case for automated decision making, the specifications in equation (9.6) is followed.

Decision making by user intervention (the optional paradigm)

In order to provide more flexibility to the occupants to interact with the energy systems while maintaining a near-optimal schedule of actions (energy-efficient routine), a slider prototype is proposed as an alternative decision-making approach. The aim of such a slider prototype is to enable the user to easily navigate from one point to the next point along the surface of the Pareto-front and to present the interdependencies of the variables to the user such that a desirable routine can be chosen by the user.

This slider prototype is obtained from the points along the estimated Pareto-front ($\mathcal{S}_{avg}^{PF}$). A global response surface is fitted using a combination of several radial basis functions using the approach in [3]. It has several bars, each one corresponding to an objective, and uses the optimization result to generate this prototype. Each of these bars (in figure **??**) are partitioned into infeasible, Pareto-optimal and non-optimal regions and the slider position is initialized to the default position as obtained from equation (9.6). The user is allowed to voluntarily navigate these sliders. However, all the sliders are co-dependent according to the fitted response surface and hence, moving one of the sliders could result in moving the other sliders into infeasible and/or non-optimal regions. When all the sliders are set at desired feasible positions (which could also be non-optimal objective values), the interpolated solution vector is presented as the recommended schedule of actions ($\mathcal{A}^{\star}$) for further analysis.

The screenshot of such a slider window from the human-computer interface (presented in the six xhapter) used at building simulation program is which shows that the occupant has chosen a feasible solution, however, a little degradation in air quality due to the movement of the sliders can lead to the setting of an unrealistic case as the desired scenario. Similar sliders are common to interactive interfaces. For example, the slider tries to choose a trade-off among comfort, cost, and wastages (ecological).

9.1.2 Optimization results and discussions

The purpose of an optimization module in the proposed framework is to ensure that a Pareto-optimal solution (better than the historical routine) has been reached which provides better indoor ambience to the occupants. Moreover, such a module allows the users to experience the occupant-building-environment interaction through the energy systems with minimal delay i.e. without undergoing the inconvenient task of simulations of all possible hypothetical routines, rather a directed exploration towards the set of best trade-offs can be performed to obtain the desirable solution. For such zero-cost human-based energy retrofit planning, performance analysis of the optimization module is necessary.

The proposed approach is implemented for the office room of Grenoble

Institute of Technology on a computer with 8 GB RAM and Intel Core i7 processor (having 2.20×10^9 s^{-1} clock speed[3]) using Python 3.4.

For addressing the many-objective minimization problem, at first, the efficacy of several optimization algorithms at solving this problem is assessed. Following this, the necessity of each objective is discussed along with the results of a random day.

Choosing a many-objective optimization algorithm

The algorithms investigated for solving the associated optimization problem, along with their specifications, are as follows:

1. In some previous works, such as [36] and [9], a weighted combination of objectives is used to transform a multi-objective optimization problem into a single-objective problem and then solved using a single-objective optimization algorithm. For the sake of comparison with such an approach, weighted combination of objectives (as in equation (9.6)) is used and the optimization problem is addressed using Simulated Annealing (SA) [65]. For defining the neighborhood in SA, the maximum ratio of values allowed to change is kept at 0.1, and the radius is allowed to be attenuated by 1 in each of the 1000 iterations. The temperature is considered to be linearly decreasing with respect to iterations. The best solution over 100 runs of SA is considered in the result.

2. Due to the popularity of genetic algorithms [90] as evolutionary optimization algorithms, the famous multi-objective version of genetic algorithm viz. Non-dominated Sorting based Genetic Algorithm-II (NSGA-II) [27] is used along with single point binary crossover (where binary tournament selection is used to choose parent vectors) and binary

[3] 10^9 s^{-1} = 1 gigahertz

mutation (where the probability of mutation is set as the inverse of the length of a solution vector).

3. Similar to the genetic algorithm, another very popular evolutionary optimization algorithm is the differential evolution. The classical version of Differential Evolution for Multi-objective Optimization (DEMO), introduced in [100], is studied for comparison. This version is indicated as DEMO'05, the scale factor for mutation is randomly generated between 0 and 2 and the crossover rate is set as 0.8. The variables in the solution vector are rounded so that the search space remains binary-valued. For incorporating elitism, it uses non-dominated sorting approach followed by crowding distance based selection (as in NSGA-II) to form the parent population of the next generation from the parent population and child population of the current generation.

4. A non-elitist but fast version of Differential Evolution for Multi-objective Optimization (DEMO), introduced in [4], is also studied in this work. For this version of DEMO (indicated as DEMO'17), the same reproduction operators are used as done in DEMO'05. The only difference is that the selection of candidates, for the next generation, is dictated by Pareto-dominance [54] followed by ranking based on equation (9.6).

5. As this particular application does not require the exact globally optimal solution, rather an approximation of the optimal solution is preferred which can be obtained speedily, hence, this work also studies Approximation-Guided Evolutionary algorithm (AGE-II) [116], which incorporates the formal notion of additive approximation. The hyper-parameter (ϵ) controls the degree of approximation and is set at 0.01, for this work. This algorithm uses the same reproduction operators as done in NSGA-II. It should be noted that AGE-II is an elitist algorithm and

maintains an external archive which stores all non-dominated solutions discovered over all the iterations.

For NSGA-II, DEMO'17, DEMO'05, and AGE-II, a maximum of 1500 iterations is used as the stopping condition. Each of these algorithms is executed 5 times and the best result is noted for performance analysis.

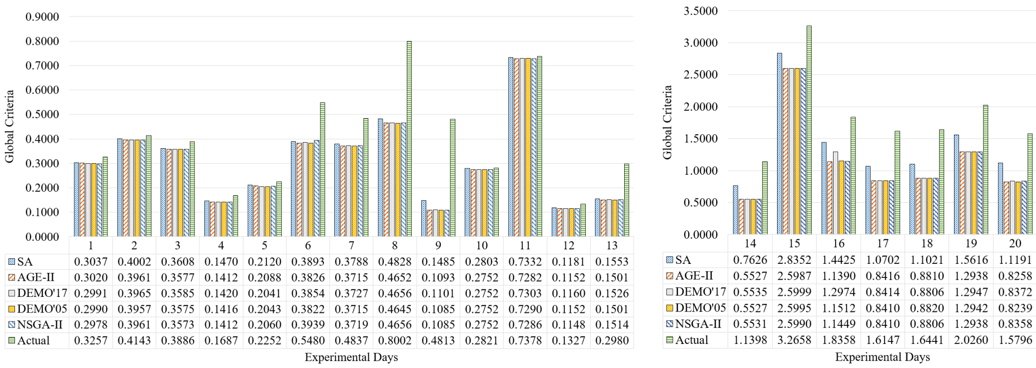
For 20 randomly sampled days, spread uniformly over the entire duration of the experiment, the optimization results are noted for the five optimization algorithms viz. SA, NSGA-II, DEMO'05, DEMO'17, and AGE-II. The automated decision making is performed to yield the preferred solution ($\mathcal{S}^{\star}_{avg}$ and $\mathcal{A}^{\star}$ from equation (9.6)).

The global criteria ($\mathcal{S}^{\star}_{avg}$) is considered as a performance indicator and lower the value of this indicator, better is the minima attained by an optimization algorithm. The results, in terms of this global criteria, are presented in figure 9.1 where along with the global criteria ($\mathcal{S}^{\star}_{avg}$) attained by simulating the optimal schedule of actions, the global criteria ($\tilde{\mathcal{S}}_{avg}$) attained by the historical (or actual) schedule of actions is also plotted for comparison.

Using the global criteria values, the difference of optimal values from the actual value is also noted in Table 9.1. Higher difference values indicate more scope of improvement has been discovered. These values assist in performing the one sample two-tailed t-test for statistical validation of the results. The p-values resulting from the t-test are noted corresponding to 95% confidence interval under the null hypothesis that the difference is insignificant i.e. mean difference is zero.

By analyzing figure 9.1 and table 9.1, the following insights are obtained:

1. In each of these cases, the optimization yields a lower value of $\mathcal{S}^{\star}_{avg}$ (other bars) than the value of $\tilde{\mathcal{S}}_{avg}$ (green horizontally striped bars). This is also indicated by the positive values of difference in global criteria and significant p-values (p-value ≥ 0.05, rejecting the null hypothesis)



(a) (b)

Figure 9.1: Comparing the performance of optimizers when global criteria with respect to actual schedule of occupant's actions is (a) less than or equal to 1, (b) greater than 1

Table 9.1: Comparison of optimizer performance

Day Number	Date	Deviation in optimal global criteria from actual (ΔS) SA	AGE-II	DEMO'17	DEMO'05	NSGA-II	Execution time (in seconds) SA	AGE-II	DEMO'17	DEMO'05	NSGA-II
1	01-Apr-2015	0.0220	0.0237	0.0266	0.0267	0.0279	61.8668	91.6743	41.1361	156.2126	101.3238
2	20-May-2015	0.0142	0.0183	0.0178	0.0187	0.0183	59.3280	97.6341	41.0231	157.3414	103.1781
3	30-Sep-2015	0.0278	0.0309	0.0301	0.0311	0.0314	60.9106	84.8661	41.0891	149.4678	100.1119
4	08-Oct-2015	0.0217	0.0275	0.0266	0.0270	0.0275	62.3862	89.8696	39.7261	149.5708	101.8488
5	03-Nov-2015	0.0132	0.0164	0.0211	0.0209	0.0192	64.7919	90.9098	43.2877	124.6023	102.4084
6	07-Dec-2015	0.1587	0.1654	0.1626	0.1658	0.1541	61.0254	86.0383	40.8998	126.9851	100.6460
7	26-Jan-2016	0.1049	0.1122	0.1110	0.1122	0.1118	61.6363	83.7820	40.5204	142.0330	101.3354
8	01-Feb-2016	0.3173	0.3349	0.3345	0.3357	0.3345	62.4322	85.8027	41.2623	140.4504	101.1520
9	17-Mar-2016	0.3329	0.3720	0.3713	0.3728	0.3728	60.5009	84.0765	41.5491	129.5074	100.3140
10	13-Apr-2016	0.0018	0.0069	0.0069	0.0069	0.0069	62.8810	85.6809	41.6365	149.3256	102.4386
11	23-May-2016	0.0046	0.0096	0.0075	0.0088	0.0092	41.6105	67.8174	29.1863	111.7215	71.4416
12	02-Jun-2016	0.0146	0.0175	0.0167	0.0175	0.0180	40.9297	67.6796	29.1433	115.1861	74.3269
13	20-Oct-2016	0.1427	0.1479	0.1454	0.1479	0.1467	62.0120	101.5243	40.8174	157.5271	100.8283
14	16-Jun-2015	0.3773	0.5871	0.5863	0.5872	0.5867	62.4832	89.1091	43.5668	146.4297	103.2391
15	07-Jul-2015	0.4306	0.6671	0.6659	0.6663	0.6668	62.8645	88.0524	41.9357	142.2077	100.6763
16	01-Sep-2015	0.3933	0.6968	0.5384	0.6846	0.6909	63.3529	94.0166	43.9191	112.7689	105.1845
17	30-Jun-2016	0.5444	0.7731	0.7732	0.7737	0.7737	42.0875	68.1006	28.9649	107.3496	73.8499
18	26-Jul-2016	0.5420	0.7631	0.7635	0.7621	0.7635	42.3909	70.1146	28.5822	100.1373	72.1790
19	31-Aug-2016	0.4644	0.7321	0.7313	0.7317	0.7321	42.2508	65.3673	29.3940	103.0661	72.9211
20	08-Sep-2016	0.4605	0.7539	0.7424	0.7557	0.7438	42.9969	63.3149	29.3030	91.1478	74.5413
Mean		**0.2194**	**0.3128**	**0.3040**	**0.3127**	**0.3118**	**56.0369**	**82.7716**	**37.8472**	**130.6519**	**93.1973**
***p*-value**		**0.000136**	**0.000306**	**0.000304**	**0.000297**	**0.000305**	–	–	–	–	–

in table 9.1. This indicates the potential of the proposed approach for energy management in buildings.

2. Under a close inspection, it can be noticed that among all the optimization results, SA yields the worst minimal global criteria (blue checkered bars are lower than the green horizontally striped bars but higher than all the other remaining bars for each of the experimental days in figure 9.1 and least deviation from actual value is seen for SA in table 9.1). This is because the objectives are indeed very conflicting in nature, as has been explained in section 9.1, however, transforming multiple objectives into a single objective leads to neglecting this conflict during optimization. Thus, single objective optimization approaches are not recommended for this kind of problems.

3. With only a few exceptions, all the remaining multi-objective optimization algorithms viz. NSGA-II, DEMO'05, DEMO'17, and AGE-II have similar bar heights in figure 9.1. When the optimal global criteria values and the difference in global criteria values, for these four algorithms, are noted, these are found to be very similar to each other. This indicates all the multi-objective optimization algorithms are capable of finding the approximation of the desired solution.

4. Besides, the minimal global criteria, the speed at which this solution is attained contributes towards the final choice of the optimizer. The speed of the optimization algorithm is considered with respect to its execution time (in seconds) as mentioned in table 9.1. It can be observed, the speed of the algorithms are in the following order: DEMO'17, SA, AGE-II, NSGA-II and then, DEMO'05.

   The higher speed of DEMO'17 can be attributed to fact that it is a non-elitist approach and hence, in each iteration, the comparison is among a smaller subset of solutions [4]. The lower speed of DEMO'05 [100] and

NSGA-II [27] is because of the non-dominated sorting of a larger subset of the population, for elitism, which is a very computationally expensive step. The higher speed of SA [65] is because it is a single-objective optimization algorithm, hence, comparing solution vectors is simpler. The intermediate speed of AGE-II [116] is because of the incorporation of archive-based elitism and the use of ϵ-grid to relax Pareto-dominance during selection.

5. During summer in France i.e. from mid-June to mid-September, the outside weather is less favorable to obtain occupant's comfort. It is observed not only by higher global criteria value resulting from actual schedule (figure 9.1b) but also from the optimal solutions achieved in these cases (experimental day number 14 to 20) which does not reach minima as close to other experimental days. Hence, this period (summer) of case study needs more actions in terms of some assistive external agents like cooling devices.

Since among the elitist multi-objective optimization algorithms, AGE-II is fastest, it is used for further analysis.

Results of a random day for analyzing the efficacy of each objective

On a random day (03 November 2015), the optimization performance in terms of hourly indoor temperature (T_{in}^k), hourly indoor CO_2 concentration (C_{in}^k), hourly window and door status (ζ_W^k and ζ_D^k) and hourly heating power consumption (P_{fuel}^k) are plotted in figure 9.2. Along with the optimal variables (green dotted curves), the historical values (blue dashed curves) of each of the parameters are also presented for reference. The following problem features can be observed from these results:

1. The performance is noted in terms of a 2-objective problem aiming at minimization of σ_{temp} and σ_{air}. Here, the data from 8 am to 8 pm

is studied as a preliminary problem. For T_{in} and C_{in} (first column of figure 9.2), it is important to note that the green dotted curve has a lower trend value than the blue dashed curve. As the optimization aims to minimize the average value over the entire day, at some hours the optimal value is greater than the historical value. Since it is an autumn day (03 November 2015), the doors and windows are closed for most of the day. However, the optimal values demonstrate that occasionally closing the doors and windows can benefit the occupant's satisfaction by regulating the indoor physical parameters. The Pareto-front and the selected final solution for this 2-objective problem are also shown to demonstrate the automated decision-making.

2. The above experiment is made into a 3-objective optimization problem by bringing the heater power consumption (P_{fuel}) into the scenario. The corresponding plots are presented in the second column of figure 9.2. It can be seen from the historical values that the heater has indeed been used by the occupants for the same context. The observations, for this case, are very similar to the 2-objective optimization problem. The optimization objective (equation (9.2)) tends to bring T_{in} between 294.15K to 296.15K. Hence, when the blue dashed curve for $T_{in} <$ 294.15K, the green dotted curve is above the blue dashed curve and vice-versa when the blue dashed curve for $T_{in} >$ 296.15K. The heater allows a more regulated control over the T_{in} than the 2-objective problem. The basis of this claim is the observation that T_{in} could be further lowered for the same time quantum with the use of the heater than without using the heater.

3. Finally, the concerned problem i.e. the 4-objective optimization problem is created by additionally aiming for the minimization of the number of changes in window and door status (δ_{WD}). The third column of figure

9.2 demonstrate the associated results. It can be observed that almost similar pattern of T_{in} and C_{in} as for the 3-objective problem can be attained with fewer changes in door and window actions. Hence, the regulation of physical parameters and, in turn, the energy management, can occur without too much interfering with the daily schedule of the occupants.

Based on these results, further analysis is performed on the 4-objective optimization problem.

Year-round results to study seasonal variations

To further investigate the seasonal variations in the optimization results, the optimization results and historical values are presented for one year (from September 2015 to August 2016) in figure 9.3 using notched box-and-whisker plots. Due to space constraint, the daily average values are provided at this webpage[4]. Hence, the four seasons appearing in this period are autumn of 2015 (Autumn'15), winter at the end of 2015 and at the beginning of 2016 (Winter'15-16), spring of 2016 (Spring'16) and summer of 2016 (Summer'16). With respect to figure 9.3, the following observations are noted:

1. Average values of T_{in} (historical and optimal) per day are noted and reflected per season in figure 9.3a. It can be observed that the seasonal average of $T^{\star}_{in}$ is higher than $\tilde{T}_{in}$ in winter and spring, and vice-versa during autumn and summer. This is because the optimization module tries to achieve the desired indoor temperature (294.15K to 296.15K).

2. Average values of C_{in} (historical and optimal) per day are noted and reflected per season in figure 9.3b. It can be observed that the seasonal average of $C^{\star}_{in}$ is lower than $\tilde{C}_{in}$ in all seasons except winter. This can

[4] `http://worksupplements.droppages.com/buildingenergy.html`

be attributed to the fact that, in winter, windows are closed most of the times to reduce φ_{out}, which, in turn, affects Q_{out} and consequently, C_{in}.

3. The variation in the average values of all the four optimization criteria is noted per season in figure 9.3c which represents the effects (simulation results) of implementing the optimal actions. From the plot, it can be noted that all the optimization criteria have approached near minimal values, except thermal dissatisfaction in summer. During high T_{out}, air cooling devices can be beneficial to enhance occupant's satisfaction [69]. Also, this high thermal dissatisfaction during summer strengthens the findings in figure 9.1b.

Although the above results demonstrate the significant advantage of following the optimal schedule of actions, yet it is difficult to convince the users to adapt to the proposed approach, especially when the underlying physical phenomena are presented in terms of graphs as in figure 9.2. A more meaningful representation and analysis of variation in occupants' actions are considered in next section.

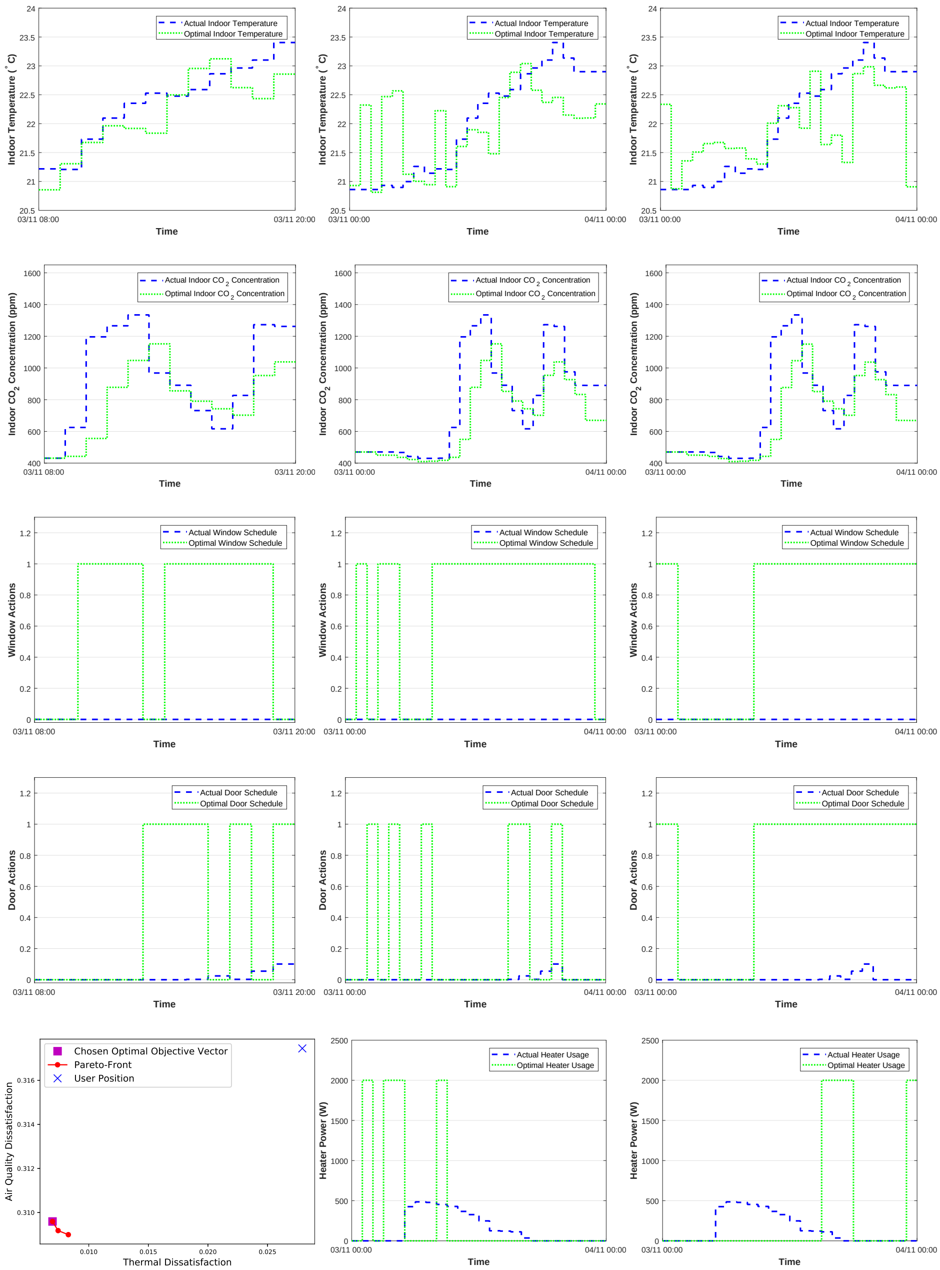
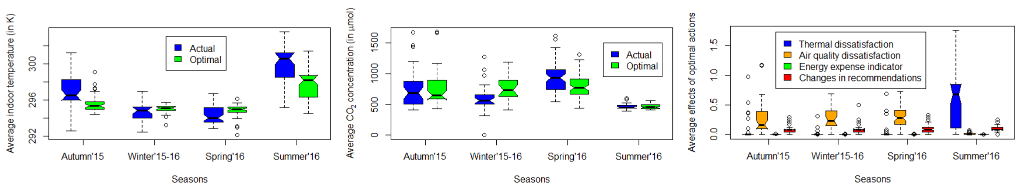


Figure 9.2: Indoor physical parameters along optimal actions for 2-objective problem with σ_{temp} and σ_{air} (first column); for 3-objective problem with σ_{temp}, σ_{air} and σ_{cost} (second column); and for 4-objective problem with σ_{temp}, σ_{air}, σ_{cost} and δ_{WD} (third column)



(a) (b) (c)

Figure 9.3: Seasonal variations in: (a) daily average of indoor temperature, (b) daily average of indoor CO_2 concentration, and (c) daily average of effects of optimal actions

Bibliography

[1] United States Environmental Protection Agency. Global greenhouse gas emissions data, 2018.

[2] JF Allen. 'maintaining knowledge about temporal intervals', cacm 26, 11, 832–843. *Google Scholar*, 1983.

[3] P. Alotto, M. Gaggero, G. Molinari, and M. Nervi. A ldquo;design of experiment rdquo; and statistical approach to enhance the ldquo;generalised response surface rdquo; method in the optimisation of multiminima problems. *IEEE Transactions on Magnetics*, 33(2):1896–1899, Mar 1997.

[4] Amr Alzouhri Alyafi, Monalisa Pal, Stéphane Ploix, Patrick Reignier, and Sanghamitra Bandyopadhyay. Differential explanations for energy management in buildings. In *IEEE Technically Sponsored SAI Computing Conference*, pages 507–516, Jul 2017.

[5] Fatima Amara, Kodjo Agbossou, Alben Cardenas, Yves Dubé, and Sousso Kelouwani. Comparison and simulation of building thermal models for effective energy management. *Smart Grid and renewable energy*, 6(04):95, 2015.

[6] Manar Amayri, Abhay Arora, Stephane Ploix, Sanghamitra Bandhyopadyay, Quoc-Dung Ngo, and Venkata Ramana Badarla. Estimating

occupancy in heterogeneous sensor environment. *Energy and Buildings*, 129:46–58, 2016.

[7] Manar Amayri, Stephane Ploix, and Sanghamitra Bandyopadhyay. Estimating occupancy in an office setting. In *Sustainable Human–Building Ecosystems*, pages 72–80, 2015.

[8] Pascal Amphoux. Vers une théorie des trois conforts. *Annuaire 90*, pages pp. 27–30, 1990.

[9] Ehsan Asadi, Manuel Gameiro Da Silva, Carlos Henggeler Antunes, and Luís Dias. Multi-objective optimization for building retrofit strategies: A model and an application. *Energy and Buildings*, 44:81–87, 2012.

[10] William Bechtel. Explanation: A mechanist alternative. *Studies in History and Philosophy of Biol and Biomed Sci*, 36(2):421–441, 2005.

[11] C. (2012) Berment, V. et Boitet.

[12] Daniel G Bobrow. *Qualitative reasoning about physical systems*, volume 1. Elsevier, 2012.

[13] Réd. (1982) Boitet C. "dse-1"— le point sur ariane-78 début 1982. contrat adi/cap-sogeti/champollion (3 vol.), geta, grenoble, janvier 1982, 616 p.

[14] Harald Brandl, Gordon Fraser, and Franz Wotawa. Qr-model based testing. In *Proceedings of the 3rd international workshop on Automation of software test*, pages 20–17. ACM, 2008.

[15] Bert Bredeweg and Paulo Salles. Mediating conceptual knowledge using qualitative reasoning. *WIT Transactions on State-of-the-art in Science and Engineering*, 34, 2009.

[16] Patrick Brézillon. Context in human-machine problem solving: A survey. *LIP*, 6(1996):029, 1996.

[17] Patrick Brézillon and J-Ch Pomerol. Contextual knowledge sharing and cooperation in intelligent assistant systems. *Le Travail Humain*, pages 223–246, 1999.

[18] Sanjay Chakraborty, N.K.Nagwani, and Lopamudra Dey. Weather forecasting using incremental k-means clustering. 2012.

[19] TY Chen. Real-time predictive supervisory operation of building thermal systems with thermal mass. *Energy and buildings*, 33(2):141–150, 2001.

[20] European Commission. Eu climate action, 2018.

[21] Enrico Costanza, Sarvapali D. Ramchurn, and Nicholas R. Jennings. Understanding domestic energy consumption through interactive visualisation: A field study. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, pages 216–225, New York, NY, USA, 2012. ACM.

[22] CSTB. Thermal regulations for new buildings in france, 2012.

[23] Wettschereck D. and Aha D.W. Weighting features. *Veloso M., Aamodt A. (eds) Case-Based Reasoning Research and Development. ICCBR 1995. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence), vol 1010. Springer, Berlin, Heidelberg*, 1995.

[24] Fred D Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, pages 319–340, 1989.

[25] Johan De Kleer and John Seely Brown. A qualitative physics based on confluences. *Artificial intelligence*, 24(1-3):7–83, 1984.

[26] Minitere de L'Environment L'energie et le mer. Chiffres clés de l'énergie - edition 2016, 2016.

[27] Kalyanmoy Deb, Samir Agrawal, Amrit Pratap, and Tanaka Meyarivan. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: Nsga-ii. *Lecture notes in computer science*, 1917:849–858, 2000.

[28] Daniel Clement Dennett. *The intentional stance*. MIT press, 1989.

[29] Anind K Dey, Gregory D Abowd, and Daniel Salber. A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human–Computer Interaction*, 16(2-4):97–166, 2001.

[30] Paul Dourish. What we talk about when we talk about context. *Personal and ubiquitous computing*, 8(1):19–30, 2004.

[31] Marek J Druzdzel. Verbal uncertainty expressions: Literature review. *Pittsburgh, PA: Carnegie Mellon University, Department of Engineering and Public Policy*, 1989.

[32] Marek J Druzdzel. *Probabilistic reasoning in decision support systems: from computation to common sense*. PhD thesis, Carnegie Institute of Technology, 1993.

[33] Marek J Druzdzel. Genie: A development environment for graphical decision-analytic models. In *Proceedings of the AMIA Symposium*, page 1206. American Medical Informatics Association, 1999.

[34] Wlodzislaw Duch and Karol Grudzi nski. Weighting and selection of features. 1999.

[35] EIA. International energy outlook 2017, 2017.

[36] Yuling Fan and Xiaohua Xia. A multi-objective optimization model for building envelope retrofit planning. *Energy Procedia*, 75:1299–1304, 2015.

[37] Lukáš Ferkl and Jan Širokỳ. Ceiling radiant cooling: Comparison of armax and subspace identification modelling methods. *Building and Environment*, 45(1):205–212, 2010.

[38] BJ Fogg. A behavioral model for persuasive design. *Proc. Persuasive 2006*.

[39] Brian J Fogg. Persuasive computers: perspectives and research directions. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 225–232. ACM Press/Addison-Wesley Publishing Co., 1998.

[40] Brian J Fogg. Persuasive technology: using computers to change what we think and do. *Ubiquity*, 2002(December):5, 2002.

[41] Kenneth D Forbus. Qualitative physics: Past, present, and future. In *Exploring artificial intelligence*, pages 239–296. Elsevier, 1988.

[42] Sylviane Gentil, Jacky Montmain, and Christophe Combastel. Combining fdi and ai approaches within causal-model-based diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(5):2207–2221, 2004.

[43] Alyssa Glass, Deborah L McGuinness, and Michael Wolverton. Toward establishing trust in adaptive agents. In *Proceedings of the 13th international conference on Intelligent user interfaces*, pages 227–236. ACM, 2008.

[44] Noah D Goodman, Tomer D Ullman, and Joshua B Tenenbaum. Learning a theory of causality. *Psychological review*, 118(1):110, 2011.

[45] Alison Gopnik and Laura Schulz. Mechanisms of theory formation in young children. *Trends in cognitive sciences*, 8(8):371–377, 2004.

[46] Arthur C Graesser and Brent A Olde. How does one know whether a person understands a device? the quality of the questions the person asks when the device breaks down. *Journal of Educational Psychology*, 95(3):524, 2003.

[47] Shirley Gregor and Izak Benbasat. Explanations from intelligent systems: Theoretical foundations and implications for practice. *MIS quarterly*, pages 497–530, 1999.

[48] Yanis Hadj Said. *Complexity in modeling in building energy management*. Theses, Université Grenoble Alpes, July 2016.

[49] Fritz Heider. *The Psychology of Interpersonal Relations*. Psychology Press, 1958.

[50] Sumi Helal, William Mann, Hicham El-Zabadani, Jeffrey King, Youssef Kaddoura, and Erwin Jansen. The gator tech smart house: A programmable pervasive space. *Computer*, 38(3):50–60, 2005.

[51] Carl G Hempel and Paul Oppenheim. Studies in the logic of explanation. *Philosophy of science*, 15(2):135–175, 1948.

[52] M Henrion. Some practical issues in constructing belief networks. uncertainty in artificial intelligence 3. 1989.

[53] Phyllis Illari. Mechanistic explanation: Integrating the ontic and epistemic. *Erkenntnis*, 78(2):237–255, 2013.

[54] Hisao Ishibuchi, Noritaka Tsukamoto, and Yusuke Nojima. Evolutionary many-objective optimization: A short review. In *IEEE congress on evolutionary computation*, pages 2419–2426, 2008.

[55] Finn V Jensen. *An introduction to Bayesian networks*, volume 210. UCL press London, 1996.

[56] Li Jiang, Da-You Liu, and Bo Yang. Smart home research. In *Machine Learning and Cybernetics, 2004. Proceedings of 2004 International Conference on*, volume 2, pages 659–663. IEEE, 2004.

[57] Daniel Kahneman, Paul Slovic, and Amos Tversky. (1982) judgment under uncertainty: Heuristics and biases. *Judgement under uncertainty: Heuristics and biases*, pages 3–20, 1982.

[58] L Karsenty and Patrick J Brezillon. Cooperative problem solving and explanation. *Expert Systems with Applications*, 8(4):445–462, 1995.

[59] Tomoya Kawakami, Naotaka Fujita, Tomoki Yoshihisa, and Masahiko Tsukamoto. An evaluation and implementation of rule-based home energy management system using the rete algorithm. *The Scientific World Journal*, 2014, 2014.

[60] Frank C. Keil. Explanation and Understanding. *Annual review of psychology*, 57:227–254, 2006.

[61] Frank C Keil. Explanation and understanding. *Annu. Rev. Psychol.*, 57:227–254, 2006.

[62] Charles Kemp, Amy Perfors, and Joshua B Tenenbaum. Learning overhypotheses with hierarchical bayesian models. *Developmental science*, 10(3):307–321, 2007.

[63] kepner tregoe. Help! my car is allergic to vanilla ice cream; a study in problem solving. the complete case, 2018.

[64] Nancy S Kim and Frank C Keil. From symptoms to causes: Diversity effects in diagnostic reasoning. *Memory & cognition*, 31(1):155–165, 2003.

[65] Scott Kirkpatrick, C Daniel Gelatt, and Mario P Vecchi. Optimization by simulated annealing. *science*, 220(4598):671–680, 1983.

[66] Johan De Kleer. Qualitative and quantitative knowledge in classical mechanics. 1975.

[67] Benjamin Kuipers. Qualitative simulation. *Artificial intelligence*, 29(3):289–338, 1986.

[68] Carmen Lacave and Francisco J Díez. A review of explanation methods for bayesian networks. *The Knowledge Engineering Review*, 17(2):107–127, 2002.

[69] Jared Langevin, Jin Wen, and Patrick L Gurian. Including occupants in building performance simulation: integration of an agent-based occupant behavior algorithm with energy plus. In *ASHRAE/IBPSAUSA Buidling Simulation Conference. Atlanta, GA*, pages 1–8, 2014.

[70] Yann Laurillau, Gaëlle Calvary, and Van Bao Nguyen. Up!: Engineering persuasive interactive. In *Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, EICS '18, pages 13:1–13:6, New York, NY, USA, 2018. ACM.

[71] Environmental Leader. Smart building energy management to top 1 bn by 2020, 2018.

[72] Shang Lei. A feature selection method based on information gain and genetic algorithm. *Computer Science and Electronics Engineering (ICCSEE), 2012 International Conference, Hangzhou, China*, 2012.

[73] K. Li, Z. Ma, D. Robinson, and J. Ma. A two-level clustering strategy for energy performance evaluation of university buildings. *4th International Conference On Building Energy, Environment*, 2018.

[74] Stéphane PLOIX Lisa SCANU, Pierre BERNAUD and Etienne WURTZ. Méthodologie pour la comparaison de structures de modèles simplifiés. France, 2016. IBPSA.

[75] Jason Liszka. Bayesian networks and causality, 2013.

[76] Tania Lombrozo. The structure and function of explanations. *Trends in cognitive sciences*, 10(10):464–470, 2006.

[77] Deborah J MacInnis, Christine Moorman, and Bernard J Jaworski. Enhancing and measuring consumers' motivation, opportunity, and ability to process brand information from ads. *The Journal of Marketing*, pages 32–53, 1991.

[78] Sebastien Mazac. *Bootstrapping sensory-motor patterns for a constructivist learning system in continuous environments based on decentralized multi-agent approach : application to ambient intelligence*. Theses, Université Claude Bernard - Lyon I, October 2015.

[79] Neil McIntyre and Mirella Lapata. Learning to tell tales: A data-driven approach to story generation. In *ACL 2009, Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 2-7 August 2009, Singapore*, pages 217–225, 2009.

[80] Robert A Melter. Some characterizations of city block distance. *Pattern recognition letters*, 6(4):235–240, 1987.

[81] Jacky Montmain and Lydie Leyval. Causal graphs for model based diagnosis. *IFAC Proceedings Volumes*, 27(5):329–337, 1994.

[82] Bonnie M Muir. Trust in automation: Part i. theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, 37(11):1905–1922, 1994.

[83] Bernard MULTON, Sylvie TURRI, Marie RUELLAN, Dominique MILLER, and Hamid BEN AHMED. Produire de l'énergie pour les appareils électroniques portables grâce aux ressources disponibles dans l'environnement humain.

[84] Gregory L Murphy and Douglas L Medin. The role of theories in conceptual coherence. *Psychological review*, 92(3):289, 1985.

[85] Kevin Patrick Murphy. Dynamic bayesian networks: Representation, inference and learning, 2002.

[86] Amro Najjar and Patrick Reignier. Constructivist ambient intelligent agent for smart environments. In *PerCom-IEEE International Conference on Pervasive Computing and Communications*, 2013.

[87] Richard E Neapolitan. *Probabilistic reasoning in expert systems: theory and algorithms*. CreateSpace Independent Publishing Platform, 2012.

[88] Raymond S Nickerson. The projective way of knowing: A useful heuristic that sometimes misleads. *Current Directions in Psychological Science*, 10(5):168–172, 2001.

[89] Monalisa Pal, Amr Alzouhri Alyafi, Stéphane Ploix, Patrick Reignier, and Sanghamitra Bandyopadhyay. Enhancing comfort of occupants

in energy buildings. In Samarjit Kar, Ujjwal Maulik, and Xiang Li, editors, *Operations Research and Optimization - FOTA, Kolkata, India, 24-26 November 2016*, pages 1–11. Springer Nature Singapore, 2016.

[90] Sokratis Papadopoulos and Elie Azar. Optimizing hvac operation in commercial buildings: a genetic algorithm multi-objective optimization framework. In *Proceedings of the 2016 Winter Simulation Conference*, pages 1725–1735. IEEE Press, 2016.

[91] Cécile L Paris. Generation and explanation: Building an explanation facility for the explainable expert systems framework. In *Natural language generation in artificial intelligence and computational linguistics*, pages 49–82. Springer, 1991.

[92] Judea Pearl. Probabilistic reasoning in intelligent systems: Networks of plausible reasoning, 1988.

[93] Nancy Pennington and Reid Hastie. Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(3):521, 1988.

[94] F. Portet, E. Reiter, A Gatt, J. Hunter, S. Sripada, Y. Freer, and C. Sykes. Automatic generation of textual summaries from neonatal intensive care data. *Artificial Intelligence*, 173(7–8):789–816, 2009.

[95] James O Prochaska and Carlo C DiClemente. *The transtheoretical approach: Crossing traditional boundaries of therapy.* Krieger Pub Co, 1994.

[96] James O Prochaska, Carlo C DiClemente, and John C Norcross. In search of how people change: applications to addictive behaviors. *American psychologist*, 47(9):1102, 1992.

[97] Ehud Reiter and Robert Dale. Building applied natural language generation systems. *Nat. Lang. Eng.*, 3(1):57–87, March 1997.

[98] Jacques Richalet. *Pratique de l'identification.* Hermes, 1991.

[99] Lance J Rips. Circular reasoning. *Cognitive Science*, 26(6):767–795, 2002.

[100] Tea Robič and Bogdan Filipič. DEMO: Differential evolution for multi-objective optimization. In *Evolutionary multi-criterion optimization*, pages 520–533. Springer, 2005.

[101] Leonid Rozenblit and Frank Keil. The misunderstood limits of folk science. *Cognitive Science*, 92:1–42, 2002.

[102] Wesley C Salmon. *Four decades of scientific explanation.* University of Pittsburgh press, 2006.

[103] Greger Sandström. *Smart Homes and User Values.* Citeseer.

[104] Usha Satish, Mark J Mendell, Krishnamurthy Shekhar, Toshifumi Hotchi, Douglas Sullivan, Siegfried Streufert, and William J Fisk. Is co2 an indoor pollutant? direct effects of low-to-moderate co2 concentrations on human decision-making performance. *Environmental health perspectives*, 120(12):1671–1677, 2012.

[105] Lisa Scanu. *Towards archetypes of self-tuned models for connected buildings.* Theses, Université Grenoble Alpes, November 2017.

[106] Lisa Scanu, Stéphane Ploix, Etienne Wurtz, and Pierre Bernaud. Model tuning approach for energy management of office and apartment settings. In *Proceedings of the 15th IBPSA Conference (Building Simulation 2017)*, pages 858–867, San Francisco, CA, USA, 2017. International Building Performance Simulation Association (IBPSA).

[107] Erik Schaffernicht and Horst-Michael Gros. Weighted mutual information for feature selection. 2011.

[108] Thomas Schütz, Markus Hans Schraven, Hassan Harb, M. C. H. W. Fuchs, and Dirk Müller. Clustering algorithms for the selection of typical demand days for the optimal design of building energy systems. 2016.

[109] Michael Scriven. Explanations, predictions, and laws. *Volume 03 (1962): Scientific explanation, space, and time*, 3(3):170–230, 1962.

[110] Henri Jacques Suermondt. Explanation in bayesian belief networks. 1992.

[111] Joshua B Tenenbaum, Charles Kemp, Thomas L Griffiths, and Noah D Goodman. How to grow a mind: Statistics, structure, and abstraction. *science*, 331(6022):1279–1285, 2011.

[112] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[113] Hugo Viot. *Modeling and instrumentation of a building and its systems to optimize energy management.* Theses, Université de Bordeaux, November 2016.

[114] GS Virk, JYM Cheung, and DL Loveday. Practical stochastic multivariable identification for buildings. *Applied mathematical modelling*, 19(10):621–636, 1995.

[115] Rudd JM Vullers, Rob Van Schaijk, Hubregt J Visser, Julien Penders, and Chris Van Hoof. Energy harvesting for autonomous wireless sensor networks. *IEEE Solid-State Circuits Magazine*, 2(2):29–38, 2010.

[116] Markus Wagner and Frank Neumann. A fast approximation-guided evolutionary multi-objective algorithm. In *Proceedings of the 15th annual conference on Genetic and evolutionary computation*, pages 687–694. ACM, 2013.

[117] William Whewell. The philosophy of the inductive sciences, founded upon their history. *Epistemology & Philosophy of Science*, 47(1):205–225, 2016.

[118] David D Woods, Emilie M Roth, and Kevin B Bennett. Explorations in joint human-machine cognitive systems. *Cognition, computing, and cooperation*, pages 123–158, 1990.

Résumé

Abstract