

5 Behavioral patterns relating to thermal comfort and energy consumption

§ 5.1 Introduction

Since the introduction of computers, the way research is performed has changed significantly. A huge amount of data can be gathered and handled by a computer, compared to the situation before these machines were commonly available to scientists and households. Every interaction with a computer system or sensor can be recorded, resulting to an abundance of data that has already surpassed the human capability to analyze and understand them. Computers are not only used for monitoring, creating and recording data but also they have become the tool to analyze these data with the use of certain automations that otherwise would make the data analysis take years.

This abundance of data has led to a new field in research, related to scientific methods and processes aiming at extracting knowledge from data in various forms [24], known as data mining. Data mining techniques have been developed to perform sequential pattern mining by processing time-ordered input streams and discover the most frequently occurring patterns [1] in applications such as healthcare, education, web usage, text mining, bioinformatics, telecommunications and other applications [17]. When data contain temporal information then they may hide additional interesting characteristics such as periodicity. A great deal of nature behaves in a periodic manner, the orbit of earth around the sun, the spinning of the planet around its axis and further on division of this periods into years, days, hours and so on. These strong periodic elements of our environment have led people to adopt periodic behavior in many aspects of their lives such as the time they wake up in the morning, the daily working hours, the weekend days off, the weekly sports practice, watching your favorite sports events or fiction series on TV every week at the same time. These periodic interactions could extend in various aspects of our lives including the relationship of people with their home thermal environment. What are the periodic elements in people's lives concerning the temperature inside their dwelling, their clothing and metabolic activity patters, their actions towards improving thermal comfort such as opening or closing windows, having a hot or cold drink or having a hot or a cold shower? These periodic elements could probably exist and are waiting to be found if a huge amount of data could be recorded and was available for analysis. Computers nowadays are powerful

enough and new mathematical methods have been developed to take advantage of this rise in computational power. Therefore, data collected by system of sensors and computers, related to the interactions of people and their residential environment could contain patterns that exhibit periodic behavior.

Recently there has been extensive research on the development of smart built environments. The goal was to reduce the energy consumption of dwellings and at the same time maintain the maximum possible comfort level for the occupants. Occupant behavior in buildings has large impact on energy consumption (space heating or cooling, ventilation demand, lighting and appliances) [2]. A number of studies have been published using stochastic models in order to model occupant presence and its interaction with space appliances and equipment.

However, all these studies were either tested in a single person office or were focused only on a specific application (occupancy [3,5,7,8], lighting [5,6,8], ventilation [4,8] etc.). Most of these works are based on the 'supervised' approach, which means that machine learning occurs by providing a set of data, and for each input value, the user provides also the output value. An (supervised) algorithm is then used to train the model and produce an inferred function, which can predict the output data when new input data is used. This method requires ground truth input data in order to be successful. For example, when talking about occupancy prediction models, the data are often based only on motion sensor readings, which could fail to detect occupants that are sitting or standing still [9]. A more complicated sensor network that includes CO₂ and humidity sensors is needed in order to have more robust occupancy and behavior detection in the residential environment than motion sensors alone [26]. The unsupervised approach on the other hand is a machine-learning task in which the user provides only input and no output data. The algorithm then is able to find the structure or relationships between the different inputs.

A smart environment, in the built environment context, is defined as an environment that is able to acquire and apply knowledge about the tenants and their physical surroundings in order to improve the tenant's experience [10] and in our case to provide insights that could lead to potential energy savings. Such an experimental network of smart environments was created during the Ecommon (Energy and Comfort Monitoring) measurement campaign, which took place in the Netherlands as part of the Monicair [11], SusLab [12] and Installaties 2020 [13] projects. Thirty-two residential dwellings were monitored for a 6-month period, from October 2014 to April 2015, which is the heating season for north Western Europe.

This study is a continuation of the work made by Ioannou et al. [14,15] under the Ecommon measurement campaign. In the above-mentioned studies, the authors used

the subjective and quantitative data related to thermal comfort to test the prediction success and the underlining assumptions of the two models widely used in this field, the PMV and the adaptive model. According to the adaptive model's main hypothesis, people are expected to perform the necessary actions, when feeling uncomfortable, that will bring them to neutral comfort sensation. Many tenants, however, had recorded "neutral" thermal sensation while the indoor temperatures were below the lower limit of the adaptive model. Furthermore, while many data points were inside the comfort band of the adaptive model, the thermal sensation votes recorded by the tenants showed comfort levels other than "neutral". Could the adaptive model be poorly estimating the tenants' adaptive capacity in relation to thermal comfort? Despite the fact that they had all kinds of options in their disposal (adjusting clothing, metabolic activity, opening or closing windows, turning up or down the thermostat, having a hot shower etc.) and the temperature was inside the comfort bandwidth, they still voted for comfort sensations other than "neutral". It could be that they exercised their adaptive options at their disposal and these were just not enough to make them feel comfortable because other parameters such as psychological ones could have a great impact. It could be the case that they did not do any of those actions. In both cases the indoor temperatures were leading the adaptive model to assume that the tenants were comfortable, having already done their adaptive actions towards thermal comfort and having "neutral" thermal sensation. But tenant's non- "neutral" feeling might lead them to take extra actions which could always come at the expense of energy consumption (especially when the tenants in the monitoring campaign answered that the economic factor plays no role in their energy spending) [14,15].

Furthermore, a statistical analysis was made with χ^2 tests between the various actions towards comfort and the thermal sensations recorded by the tenants during the monitoring campaign in order to find out which of these actions took place habitually and which were aimed towards improving thermal comfort. For example, the indoor temperature during the morning hours in some dwellings was above 20 °C, however, tenants were waking up and as a first thing they were turning up the thermostat. Moreover, other habitual actions, such as having a hot shower and opening the window, were found to be unrelated to thermal comfort and related to increased energy consumption.

The aim of this paper is to go a step further in this direction. Repetitive behavioral actions in sensor rich environments, such as the dwellings of the Ecommon measurement campaign, can be observed and categorized into patterns through data mining techniques. These discoveries could form the basis of a model of tenant behavior that could lead to a self-learning automation strategy [16] or better occupancy data to be used for better predictions of building simulation software such as Energy+ or ESP-r and others.

§ 5.2 Study design

Heierman et al. [1] described a sequential pattern mining approach that was borrowed from economics [18] and applied in the context of the built environment. An example of a sequential pattern mining application in economics is used by major supermarket chains. These supermarkets monitor the purchases of their clients (usually by a discount card in which supermarkets store information) and by applying pattern mining they try to find at a specific time of the day, which are the purchase patterns of the customer. For example, at 13:00 when the customer A is buying cheese it is most likely that he will also buy bread and orange juice. Specific patterns can be defined for the various times of the day. The same customer during the early morning hours could have a specific purchase pattern, buying for example croissants and orange juice while during the evening hours he could be buying vegetables and chicken. In the context of the built environment, the customer A can be substituted by a specific dwelling. The products that the customer can buy can be substituted by quantitative data like specific ranges in temperature (for example $18\text{ }^{\circ}\text{C} < T_{in} < 20\text{ }^{\circ}\text{C}$ or $T_{in} > 20\text{ }^{\circ}\text{C}$) or by subjective data (clothing and metabolic activity levels and actions such as opening or closing a window, having a hot shower or a hot drink).

In this study, real time data obtained by a seasonal monitoring campaign on the built environment will be implemented on the above-mentioned methodology in order to gain insights in the occupant behavior related to energy consumption of the residential sector. The main aim of this study is to demonstrate if such a pattern recognition algorithm is suitable for discovering meaningful patterns of occupancy behavior. Furthermore, this study will try to explore how these patterns can be used to improve the energy simulations for the prediction of energy consumption in the built environment.

§ 5.2.1 Research Questions and goals

The research questions and sub-questions are formulated as follows:

- 1 Can we implement an unsupervised algorithm as a data driven model for the prediction of occupant behavior related to energy consumption and thermal comfort in order to:

- discover the most frequently recorded thermal sensations, actions towards thermal comfort, and metabolic activity and clothing levels based on the tenants' recorded data?
 - discover the most frequent occurring sequences among the above mentioned items?
 - discover if there are different patterns of behavior at different times of the day?
- 2 Estimate how building energy simulations can be improved by this methodology.

§ 5.2.2 Ecommon Campaign set-up

Detailed information on the Ecommon campaign set-up, the data acquisition set, and the subjective and quantitative data gathered during the campaign can be found in the previous chapter of this thesis.

The dwellings that participated in the measurement campaign were part of the Dutch social housing stock which represents about one third of the total residential units and it is quite representative of the residential stock as a whole [27]. The sample was divided into energy A/B-rated and F-rated dwellings (Ioannou and Itard, 2017 [14]) and the final sample of the dwellings is described in Table 5.1. Finally, only seventeen dwellings were included in the analysis due to data limitations.

TABLE 5.1 Dwellings participating in the Ecommon campaign

	NO.	ENERGY RATING	HEATING SYSTEM	NO. OF OCCUPANTS	AVERAGE AGE OF HOUSEHOLD	NO. OF DATA POINTS	
						Morning hours	Evening hours
	W004	A	Heat pump	2	67	135	167
	W005	A	Condensing gas boiler	1	92	109	61
	W006	A	Condensing gas boiler	2	77	166	157
	W010	A	Condensing gas boiler	2	29	96	80
	W016	B	Condensing gas boiler	2	70	173	131
	W032	B	Condensing gas boiler	3	39	8	16
Total A/B dwellings	-	-	-	2	62.33	687	612

>>>

TABLE 5.1 Dwellings participating in the Ecommon campaign

	NO.	ENERGY RATING	HEATING SYSTEM	NO. OF OCCUPANTS	AVERAGE AGE OF HOUSEHOLD	NO. OF DATA POINTS	
						Morning hours	Evening hours
	W012	F	Condensing gas boiler	4	40.5	295	482
	W013	F	Condensing gas boiler	3	53.3	291	332
	W014	F	Gas stove	1	83	35	26
	W020	F	Condensing gas boiler	2	74	323	258
	W021	F	Condensing gas boiler	2	73	118	273
	W022	F	Condensing gas boiler	2	64	171	301
	W024	F	Condensing gas boiler	1	72	89	105
	W025	F	Gas stove	3	43	67	70
	W026	F	Condensing gas boiler	4	21	65	85
	W028	F	Condensing gas boiler	2	72	174	190
	W031	F	Condensing gas boiler	3	43	958	1924
Total F dwellings	-	-	-	2.5	58	2586	4046

§ 5.3 Sequential Pattern Mining

Sequential pattern mining methods have applications in many fields. A very common goal when using sequential mining is the discovery of the most frequent patterns [18,19]. The more frequent an event, the more important it is and more likely to be a pattern. During the analysis of time-stamped data it is important to know if event (a), event (b) and event (c) occurs frequently but it is more intriguing to know how often the event (a, b), (a, c) or the event (a, b, c) occurs. Furthermore, knowledge on the most frequent combinations of events over time, adds even more value to the analysis.

In market research, this would mean not only knowledge on which are the most common product combinations that a customer buys in his visits to the shop, but also knowing in which part of the day these occur. Customers usually buy different things in the morning and different ones in the evening and in that way shops can create tailor made marketing strategies to increase sales. In the context of the built environment this would mean that instead of tracing combinations of events that might occur in a dwelling in a whole day (which could have limited use in terms of improving thermal comfort and reducing energy consumption), now we can target specific hours and see the behavior of tenants in different periods of the day.

The algorithm that was used for the mining of sequential patterns in this study is the Generalized Sequential Pattern (GSP) algorithm [21], which is an enhanced version of the *a priori* algorithm suggested by Agrawal and Shrikant [20]. The methodology for the application of the specific technique in the context of the built environment has been described by Heierman et al. [1] but it lacked any experimental demonstration. The Ecommon campaign provided enough built environment related data that could be implemented in the above-mentioned methodology.

Input parameters

The time parameter and the customer id are inputs to the algorithm. With this pair of parameters, the algorithm is generating a sequence per customer containing every transaction made in a specific time. Then the algorithm searches sequential patterns such as: if customer A bought the item (a) and item (b) in a transaction, he bought item (c) in the next one.

Another input parameter is the *minimal support*, which describes how many customers must support a pattern in order for the algorithm to regard it as frequent. It takes values between 0 and 1 with 1 being the 100% of the customers. If we set for example the minimal support to 0.9 the algorithm will *prune* all the patterns that are supported by less than 90% of the customers.

Furthermore, three remaining input parameters are defining how transactions are handled. These are the *min-gap*, the *max-gap*, and the *window-size*. The *window-size* defines the period within successive transactions could be considered as a single transaction. For example, if a customer bought some products (a, b, c) but forgot to buy the product (d) and comes back after 10 minutes to buy this remaining product then the question is: will this transaction be treated as a completely new one or it will be added to the previous one? In order to avoid this issue the *window-size* determines how long a subsequent transaction is treated as the same transaction. In the above example if the *window-size* is larger than 10 minutes then buying the product (d) will be treated as part of his initial transaction when he bought (a, b, c).

The *max-gap* parameter is used in order to filter out large gaps in data sequences. For example, a customer bought the product (a) and despite that he is within the specified *window* there is a very large gap between buying the product (b) which is his new transaction. For a business owner this huge gap, even if it is inside the *window size*, might still make the customer uninteresting. Therefore, this is an extra tool of the GSP algorithm when seeking supported sequences. The *max-gap* parameter causes sequences not to support a pattern if the transactions containing this pattern are

time-wise too widely separated. The same applies for the *min-gap* parameter for the sequences that belong to transactions that time wise appear too near.

The concepts of the *window-size*, *min-gap* and *max-gap* parameters were the most important upgrades of the apriori algorithm, introduced by Agrawal and Shrikant [18], and led to the GSP algorithm [21]. These concepts helped to overcome important weaknesses of the apriori algorithm such as the absence of time constraints and the rigid definition of a transaction. The apriori algorithm has no time constraint, which means that the data source is a time ordered input sequence with no natural points that indicate the start or stop of the pattern. Furthermore, the user cannot specify a minimum or a maximum time gap for two adjacent elements of a sequential pattern. For example, if we were applying the apriori algorithm in the transactions of a library where a person borrowed the book (a) and then he borrowed another one after three years the algorithm would still show (a, b) as a potential pattern if the window size was three years. However, such a pattern has such a major gap between the transactions that it does not really add substantial knowledge to the library concerning the borrowing patterns of people. Setting the minimum or maximum gap into, for example, three months will automatically prune all the patterns that are not supported from this time gap and are not of interest to the library.

The rigid definition of the transactions as mentioned above is related to the *window-size*. This parameter sets the time window within successive transactions to be treated as a single transaction. For example, a person that borrows book (a) from a library, book (b) next week and book (c) the week after. If the user sets the *window-size* to three weeks then the supported pattern for that person would be (a, b, c). If the window size was two weeks then the supported patterns would be (a, b) and (c). This concept adds greatly to the flexibility of the analysis and offers much more options to the user that is mining for sequential patterns.

§ 5.3.1 Sequential pattern mining in the context of the built environment

In order to make use of an algorithm developed for the retail industry in the context of the built environment first all the input parameters have to be defined in the respected context. Furthermore, the data have to be transformed into the right format in order to be handled by the algorithm.

Input data

In the retail context the customer buys various products (transactions) in specific hours and based on his frequent combinations transaction patterns are mined. In our case, the transactions are called *events* and our customers are the people of the seventeen, dwellings that participated in the monitoring campaign. The various 'products' that our 'customer' (dwelling) can 'buy' are temperature range, recorded thermal sensation, actions towards thermal comfort, clothing, and metabolic activity levels.

Temperature Range: Houses of A/B and F label have usually a temperature range from 18-24°C which for the purposes of the pattern mining was broken down into bins of 2 °C (18 °C – 20 °C, 20 °C – 22 °C and 22 °C – 24 °C) at a given time.

- Recorded thermal sensation: is the vote casted by the occupant according to his thermal sensation at a given time of the day. It can be distinguished into 'cold', 'a bit cool', 'neutral', 'a bit warm' and 'hot'.
- Actions towards thermal comfort: Several actions that the occupants could choose towards the improvement of their thermal comfort were predefined in the comfort logbook. The options were opening or closing a window, having a hot or cold drink, put on or put off clothes, turning the thermostat up or down, having a warm or cold shower.
- Clothing: Tenants could choose from a set of predefined clothing items, which were closest to the clothing ensemble that he/she was wearing at a specific moment. The options were sleeveless t-shirt, t-shirt, knit sport shirt, long sleeved sweatshirt, jacket, jacket and hood (Table 5.2).
- Metabolic activity: occupants could also choose from a set of predefined metabolic activity levels. These levels were lying/sleeping, sitting relaxed, light deskwork, walking, jogging, running (Table 5.2).

All the above answers were given by the occupants every time bearing in mind the last 30 minutes.

All the input data for the GSP algorithm have to be binominal (nominal with two possible values, true or false). This means that the data, quantitative and subjective, had to be properly transformed to be compatible with the GSP algorithm input requirements. As already mentioned in section 2.2, the quantitative data (temperature, humidity, and CO₂) are real numbers obtained by a set of sensors with a 5-minute interval for a period of six months between October and April. For the purposes of this study, the temperature was the quantitative measurement that was used in the GSP

calculations. In order to transform the temperature into binominal data the following process took place: the 5-minute interval data were aggregated into hourly values for the whole period of two weeks and then three bins of temperatures were defined ($18 < T < 20$, $20 < T < 22$, $T > 22$). If the temperature in a specific hour was, for example, between 18 °C and 20 °C then the $18 < T < 20$ bin would take the value TRUE (for this specific hour) and the rest of the bins would take the value FALSE. The procedure is repeated until all the hourly values under the four temperature bins are transformed into TRUE or FALSE. The reasons for the hourly aggregation of the data were that the previous research of the authors [11,14,15] was based on hourly aggregation of the data due to their large volume. Furthermore, the hourly time-step is a very common time-step during building simulations and one of the major goals of the Ecommon, Monicair and Installaties2020 projects was the improvement of the prediction quality of the simulation software for the built environment. Therefore, for consistency between our goals and results so far we chose to use the hourly aggregation of the data also in this study. Furthermore, only the data that were accompanied by recorded motion data were used for the analysis in this study.

The subjective data were transformed in similar way with the difference that the bins in this case were the subjective data themselves. Thermal sensations, actions towards thermal comfort, clothing, and activity level are categories that can take binominal values for each hour of the day. For example, if a tenant has recorded that he feels 'neutral' within the 5-minute interval between 13:30 and 13:35 then for the 13th hour the value under 'neutral' bin would be TRUE while the value under all other thermal sensations would be FALSE. The same applies for the clothing, activity levels, and actions towards thermal comfort. If within the 5-minute interval between 13:30 and 13:35 of a day a tenant recorded that he wears 't-shirt' and is 'sitting relaxed' then the value under the 't-shirt' and 'sitting relaxed' bins for the 13th hour of that day would be TRUE and all the other clothing and metabolic activity options would take the value FALSE. Also, if during the 5-minute interval of an hour an occupant recorded that he has opened the window, or turned the thermostat on then at that specific hour the values of 'open window' and 'thermostat up' would be TRUE and all the rest of the actions would be false.

One limitation of this approach was, as mentioned already, that tenants were instructed to fill in the subjective data based on what they did the previous half an hour. The recording of the thermal sensation is not affected by this directive, when an occupant recorded that he felt 'neutral', 'a bit cool' or 'cool' he was recording his instantaneous thermal feeling. However, for the rest of the subjective data such as actions towards thermal comfort, clothing, and activity levels recorded data at the 13:15 hours could mean that some of these actions such as 'close window' or 'open window' could have occurred before 13:00 hours. For the clothing it is more likely that

tenants recorded what they were wearing at that exact moment with the exception of 'jacket' which indicated most of the times that people were outside and came home with in the last half hour. Nevertheless, the actions towards thermal comfort could have a delay up to half an hour. The general assumption for the purposes of this study was that during the hourly aggregation when an action, clothing or metabolic activity appeared within a specific hour's 5-minute interval then it was eventually assigned in this hour. The reason for this was that we had no way to determine the exact time an action, clothing or activity levels took place from the time it was recorded and the previous half hour. This problem could have been even more evident if we had not aggregated the data into hourly values. As already mentioned, prior research has taken place in hourly values and hourly values is a very common time step for simulation software. With hourly aggregation every action, clothing and metabolic activity recorded with timestamp in the second half hour (for example after 13:30) it had most chances to have occurred within this hour rather than before 13:00.

Finally, for the analysis not all the hours of the day were used partly because that would require a very big data file and slow computational time and partly because not all the hours of the day are of the same importance. As already mentioned only the data points with motion were kept for the analysis. Further filtering removed all the data points that had no subjective data recorded. Hourly data of thermal sensation, actions towards thermal comfort, clothing and metabolic activity that had only FALSE values were removed from the analysis. Each hourly value in order to be used for further analysis should have at least one TRUE value in the subjective parameters.

From occupant behavior related to thermal comfort point of view the most interesting hours of the day are the early morning hours when people wake up and the early evening hours when people return from work. In that sense, the morning hours between 7-9 a.m., for each day of the two weeks that occupants were given the comfort dial, were chosen for the morning analysis and the 5-7 p.m. were chosen for the evening analysis. In Table 5.2, we can see a data set example with all the necessary transformations that was used by the GS algorithm for the purposes of this study.

TABLE 5.2 Example of input file for (morning hours) sequential pattern mining with the use of the GSP algorithm in the context of residential built environment

Customer id (dwelling number)	DAY	TimeStamp (hour)	TEMPERATURE RANGE			THERMAL SENSATION			ACTIONS TOWARDS THERMAL COMFORT			CLOTHING			METABOLIC ACTIVITY		
			T<18	18<T>20	cold	A bit cool	Open window	Thermostat up	t-shirt	jacket	Sitting relaxed	Light desk work
5	1	7	TRUE	FALSE	...	TRUE	FALSE	...	FALSE	TRUE	...	TRUE	FALSE	...	FALSE	FALSE	...
5	2	7	FALSE	TRUE	...	FALSE	FALSE	...	FALSE	TRUE	...	FALSE	FALSE	...	FALSE	FALSE	...
5	3	7	TRUE	FALSE	...	FALSE	FALSE	...	FALSE	TRUE	...	TRUE	FALSE	...	FALSE	TRUE	...
5
5	1	8	TRUE	FALSE	...	FALSE	FALSE	...	FALSE	FALSE	...	FALSE	TRUE	...	TRUE	TRUE	...
5	2	8	FALSE	TRUE	...	FALSE	FALSE	...	FALSE	FALSE	...	FALSE	FALSE	...	TRUE	TRUE	...
5	3	8	FALSE	TRUE	...	FALSE	FALSE	...	FALSE	FALSE	...	FALSE	FALSE	...	TRUE	TRUE	...
5
5	1	9	FALSE	FALSE	...	FALSE	TRUE	...	TRUE	FALSE	...	FALSE	TRUE	...	TRUE	TRUE	...
5	2	9	TRUE	FALSE	...	FALSE	FALSE	...	FALSE	TRUE	...	FALSE	FALSE	...	TRUE	TRUE	...
5	3	9	FALSE	TRUE	...	FALSE	FALSE	...	FALSE	FALSE	...	TRUE	FALSE	...	FALSE	FALSE	...
..
8	1	7	FALSE	TRUE	...	FALSE	FALSE	...	TRUE	FALSE	...	FALSE	FALSE	...	TRUE	FALSE	...
8	2	7	TRUE	FALSE	...	TRUE	FALSE	...	FALSE	FALSE	...	FALSE	FALSE	...	FALSE	FALSE	...
8	3	7	TRUE	FALSE	...	TRUE	TRUE	...	TRUE	FALSE	...	FALSE	FALSE	...	FALSE	FALSE	...
8
8	1	8	TRUE	FALSE	...	FALSE	FALSE	...	FALSE	FALSE	...	FALSE	TRUE	...	TRUE	TRUE	...
8	2	8	FALSE	TRUE	...	FALSE	FALSE	...	FALSE	FALSE	...	FALSE	FALSE	...	TRUE	TRUE	...
8	3	8	FALSE	TRUE	...	FALSE	FALSE	...	FALSE	FALSE	...	FALSE	FALSE	...	TRUE	TRUE	...
8
8	1	9	FALSE	FALSE	...	FALSE	TRUE	...	TRUE	FALSE	...	FALSE	TRUE	...	TRUE	TRUE	...
8	2	9	TRUE	FALSE	...	FALSE	FALSE	...	FALSE	TRUE	...	FALSE	FALSE	...	TRUE	TRUE	...
8	3	9	FALSE	TRUE	...	FALSE	FALSE	...	FALSE	FALSE	...	TRUE	FALSE	...	FALSE	FALSE	...

The customer id, as mentioned already, denotes the dwelling under monitoring, the timestamp shows the hour under consideration (e.g. 7 means the 7th hour of the day between 6 a.m. and 7 a.m.) and the rest of the columns show the quantitative and subjective parameters that have been transformed into binominal values for the GSP algorithm simulation. In the end, there is one input string per dwelling per day per

timestamp. Temperature range and thermal sensation can have only one value that can be true for each timestamp while for the rest of the parameters more than one is possible. Furthermore, in Table 5.3 we can see the taxonomy that was used for this analysis. The analysis took place for the A/B and F dwellings for the morning and evening hours respectively.

Input Parameters

The *Customer-id* is the first input parameter. Originally, this would be the customer of a retailer as already mentioned. For the purposes of this study the customers are the seventeen respondents of each of the seventeen dwellings that were monitored during the campaign.

The *timestamp* would be the time that a retail customer would make a transaction. In our case, the quantitative data that were gathered by the wireless sensors had a granularity of 5 minutes. The data were aggregated into hourly values and so the *timestamp* could get a value between one and twenty-four with one being the first hour of the day between 00:00 and 1:00 am and 24 being the last hour of the day between 23:00 pm and 00:00.

The *minimal-support* that was used for our analysis was adjusted for each simulation until we were able to find the highest support between dwellings that was giving meaningful patterns. We started with 0.9 (which means that 90% of the dwellings support a pattern) and run one simulation each time reducing the minimal support by 0.1 at a time until meaningful patterns were revealed.

The *window-size* was assumed zero, which means that the three hours of the morning (7-9 a.m.) period and evening period (5-7 p.m.) were treated as a single time window. The reason for this choice was that for the purposes of this study we were not interested in what is happening in each hour specifically but for the morning and evening periods as a whole.

The *min-gap* and *max-gap* values were assumed to have a value of 1. The reason for this was again that we wanted to find frequent patterns in an hourly basis. By setting the *min-gap* and *max-gap* to one, we assure that all frequent patterns will be contained in the hourly basis that we have been aiming.

§ 5.3.2 Building simulations

In order to demonstrate how the sequential pattern recognition methodology can improve the energy consumption calculations for the built environment, we had to perform simulations with a whole building simulation software (Energy+). The dwellings that participated in the measurement campaign had various typologies and it was not possible to perform exact energy simulations for each one of those dwellings. However, we had abundance of data concerning the daily temperature profiles for each type of room of these dwellings, their heating system, the insulation level of their windows, and their walls (assumed from the energy label of each dwelling and the year of construction), the number of people and their occupancy profiles (derived from the motion sensors). Therefore, we used the Delft University of Technology Concept House [23] as the reference building in order to perform the simulations for the dwellings that participated in the measurement campaign. The typology of the Concept house and the dwellings was not the same, however, all other aspects of the simulation (heating system, U values for walls and windows, occupancy schedules, hourly temperature profiles for each type of room, number of people) were based on realistic data gathered during the campaign. Some of the simulation parameters were adjusted to the energy label and age of the dwellings (such as infiltration and ventilation) and others such as electricity consumption for lighting and appliances were assumed the same for all dwellings.

The heating control for each dwelling was simulated with three different ways. First, the heating set point temperature was corresponding to the indoor air temperature, followed by the indoor operative temperature and finally the PMV comfort level. The indoor temperatures for each room of each dwelling were provided by the measurement campaign's data while the PMV was set to be between the comfort levels of -0.5 and +0.5.

§ 5.4 Results

Sections 4.1 until 4.5 present the temperatures, recorded thermal sensation, actions towards thermal comfort, clothing, and activity levels for the data points that were used in the GSP analysis. Section 4.6 shows the results of the GSP analysis and 4.7 the results of the Energy+ simulations.

§ 5.4.1 Temperature

Figures 5.1 and 5.2 display the morning and evening temperatures of all dwellings for the total data points that were used for the simulations with GSP.

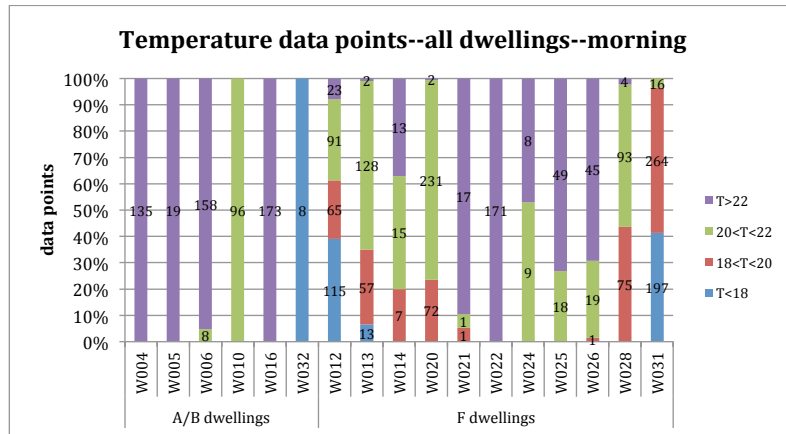


FIGURE 5.1 Morning temperatures of all dwellings for the total data points used in GSP analysis

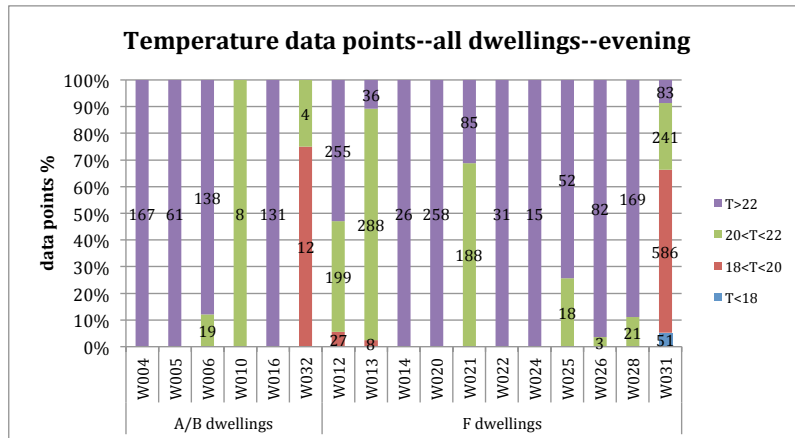


FIGURE 5.2 Evening temperatures of all dwellings for the total data points used in GSP analysis

For A/B labelled dwellings, Figure 5.1, all temperatures during the morning hours (7-9 a.m.) were above 20 °C and four out five dwellings had temperatures above 22 °C. For F dwellings, the majority of morning temperatures are above 20 °C, however, significant increase is observed in temperatures below 18 °C or between 18 °C and 20 °C. The thermal envelope of A/B dwellings could have played a significant role in this respect apart from potential occupant behavior.

For the A/B dwellings during evening hours, Figure 5.2, the temperatures of 95% of the data points were above 22 °C and the rest between 20 °C and 22°C (dwelling W010). In terms of temperature there seem to be no great differences between morning and evening hours for the A/B label dwellings. The majority of temperatures for the F labeled dwellings, approximately 75% of the data points, were above 20 °C. Compared to the morning hours there is a significant increase (more than double) in the percentage of temperatures above 22 °C and a decrease in temperatures below 20 °C, Figure 5.3. This shows clearly that occupants prefer their dwellings to be warmer in the evening than in the morning hours. In A/B labeled dwellings there is an increase in temperatures above 22 °C and a decrease in temperatures between 20 °C and 22 °C. Therefore, A/B and F label dwellings are warmer in the evening hours than in the morning hours.

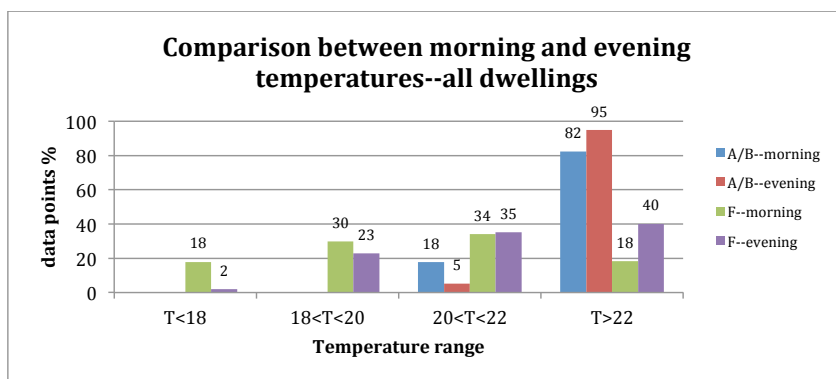


FIGURE 5.3 Comparison between morning and evening temperatures of all dwellings for the total data points used in GSP analysis

§ 5.4.2 Reported thermal sensation

Figures 5.4 and 5.5 display the total amount of reported thermal sensation scores for the data points during used for the GSP simulation.

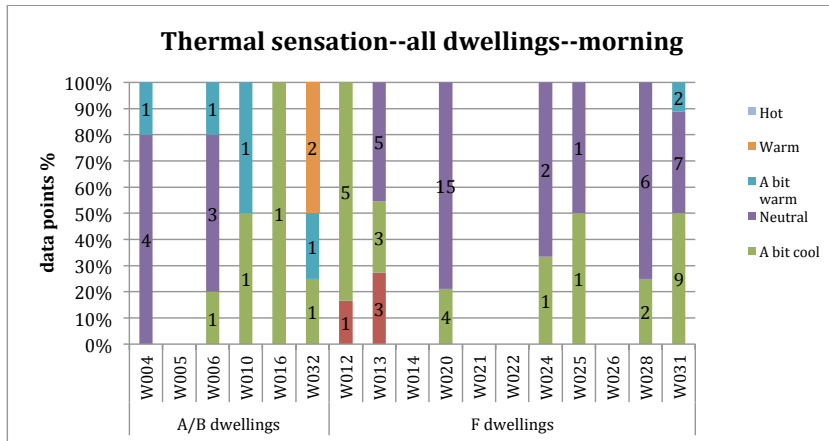


FIGURE 5.4 Morning thermal sensation scores of all dwellings for the total data points used in GSP analysis

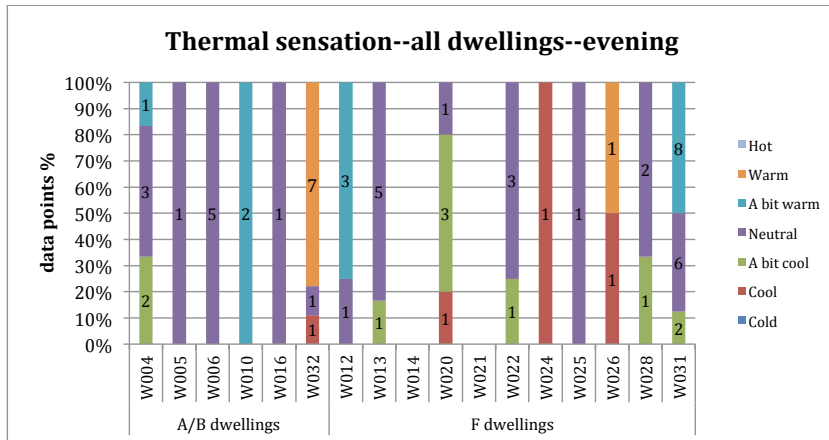


FIGURE 5.5 Evening thermal sensation scores of all dwellings for the total data points used in GSP analysis

The reported thermal sensation for the morning hours are not enough to draw conclusions, however, it is still surprising that despite the high indoor temperatures, occupants of A/B dwellings recorded thermal sensations such as 'a bit cool'. For the F labeled dwellings there were more thermal sensations reported and the majority of those were 'neutral' despite the lower indoor temperatures.

On the one hand, this could be a result of the occupants' difficulty in discriminating between the various thermal sensations [14]. The seven-point thermal sensation scale, developed in climate chambers, provides no guarantee that a specific thermal comfort level reported by a Dutch occupant corresponds to the PMV scale. Furthermore, studies have found that people's thermal sensations vary between winter and summer, from individual to individual, and are dependent on race, climate, habits and customs [29,30,31]. On the other hand, this could as well be a sign of the effect of psychological expectations. Adaptation is defined as the gradual lessening of the occupants' response to repeated environmental stimulation and can be behavioral, physiological and psychological [28]. The majority of the thermal sensations recorded in this measurement campaign were between -1 (a bit cool) and +1 (a bit warm). Analysis of these data in a prior study showed that the PMV model predicted well the thermal comfort of the occupants for thermal sensations between -1 and +1 while the prediction was getting less accurate approaching -3 or +3 [14]. These dwellings are the personal space of the occupants, a place they always try to keep a comfortable as possible, and comfort is part of what people associate with the notion of home. Occupants of the F dwellings may be aware of the lesser thermal capabilities of their homes and used to the lower indoor temperatures of their dwellings and have adapted to these conditions. If this is true, then despite the fact that these people might have lowered their thermal comfort standards, it is beneficial for the environment and energy efficiency of the housing sector because occupants could have just been using more energy in order to increase their comfort instead of adapting. All occupants in this campaign said they have no problem paying their energy bills, which they found easy to pay, despite the fact that their income ranged between half and one and a half time the Dutch median [32].

The comfort votes of the A/B dwellings during evening hours have shifted to more 'neutral' and 'a bit warm', which is logical based on the indoor temperatures. For F labeled dwellings the effect of increased temperatures during evening hours does not seem to be translated into more comfortable thermal sensation votes although still the majority of thermal sensations are between 'a bit cool' and 'a bit warm'. However, the amount of data is not sufficient to draw concrete conclusions.

§ 5.4.3 Actions towards thermal comfort

Figures 5.6 and 5.7 display the actions towards thermal comfort for the morning and evening hours used for the GSP simulation. For the morning hours, the occupants of the F labelled dwellings recorded having a 'hot drink', having a 'warm shower' and 'thermostat up' as the most common actions which seem intuitively sensible given the lower temperatures of their dwellings. These actions seem to be genuinely performed in order to improve thermal comfort. The occupants of the A/B labelled dwellings, however, have used various actions in a more erratic way. For example, W004 had morning temperatures above 22 °C for the whole period of analysis and the tenants still recorded having a warm shower and a warm drink every morning while feeling 'neutral'. Obviously, these actions in this particular case are not related to thermal comfort. Dwelling W006, with similar indoor temperatures as W004, recorded having a 'hot drink' and even turning the 'thermostat up' while thermal sensations were mainly 'neutral'. This occupant behavior could be led by behavioral reasons and could have an impact in energy consumption of a dwelling with no significant benefit to indoor comfort.

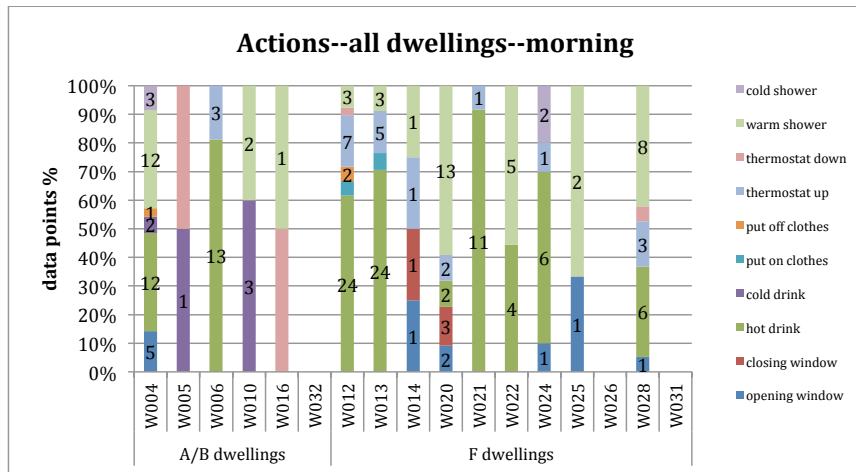


FIGURE 5.6 Morning actions toward thermal comfort scores of all dwellings for the total data points used in GSP analysis

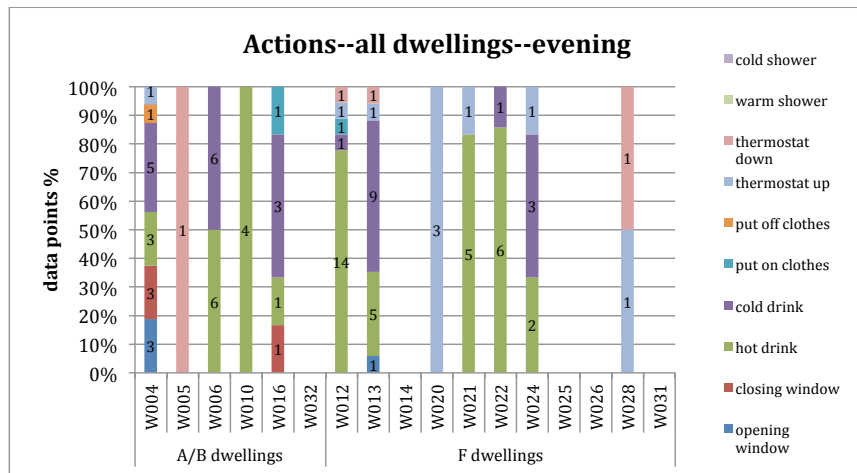


FIGURE 5.7 Evening actions toward thermal comfort scores of all dwellings for the total data points used in GSP analysis

During the evening hours, all F label dwellings that recorded ‘thermostat up’ had temperatures above 20 °C. However, only dwellings W020 and W024 had recorded majority of thermal sensations ‘a bit cool’ or ‘cool’, which could explain the action of thermostat up. All other F labelled dwellings had temperatures above 20 °C and the majority of thermal sensations were ‘neutral’ followed by ‘a bit cool’, to a lesser extent, while dwelling W012 even had thermal sensations of ‘a bit warm’. Regardless of the recorded thermal sensations, the level of indoor temperatures is very high to substantiate an action such as ‘thermostat up’, which affects energy consumption. Compared to the morning actions, for F labelled dwellings, ‘cold drink’ has been substituted with ‘warm shower’. This action in A/B labelled dwellings has substituted ‘closing the window’.

§ 5.4.4 Clothing

Figures 5.8 and 5.9 display the clothing levels for the morning and evening hours used for the GSP simulation.

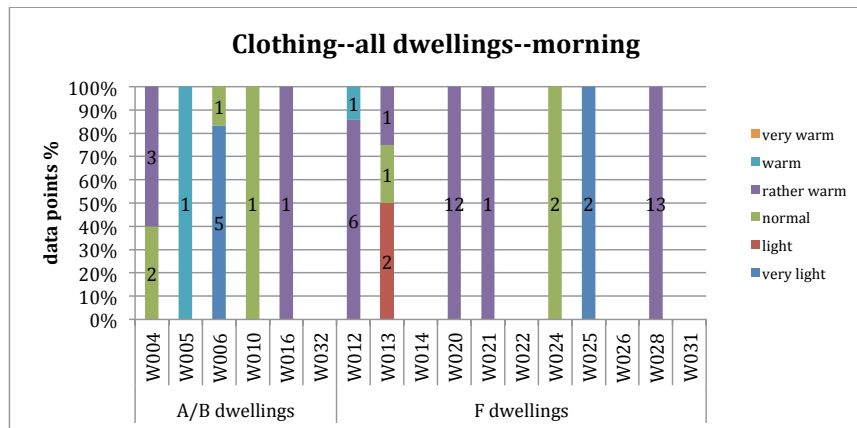


FIGURE 5.8 Morning clothing scores of all dwellings for the total data points used in GSP analysis

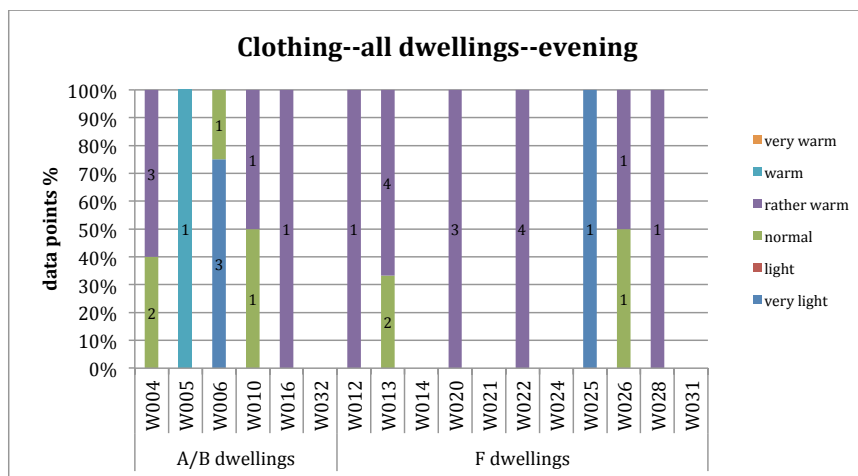


FIGURE 5.9 Evening clothing scores of all dwellings for the total data points used in GSP analysis

During the morning hours, for the F labeled dwellings, we see the majority of clothing being rather warm 'long sleeved sweat shirt'. Take dwellings W020 and W028, for example. The majority of hours between 7-9 a.m. have temperatures between $20^{\circ}\text{C} < T < 22^{\circ}\text{C}$ and the occupants mainly feel 'neutral' and a few times 'a bit cool'. The seemingly consolidated 'long sleeved sweat shirt' clothing pattern for F labeled dwellings could be part of the psychological adjustment mentioned earlier. The worst (compared to A/B dwellings) thermal conditions in these dwellings are compensated

by a higher clothing level which is a good practice concerning energy conservation. As we can see in Figure 5.10, from the 41 data points on actions towards thermal comfort recorded for W020 and W028, only 5 times there was an increase in thermostat levels during the morning hours. Occupants have adjusted themselves in order to feel neutral by means of clothing and other actions such as 'hot drink' or 'warm shower'. Temperature conditions in A/B dwellings are always above 22 °C, which allows for a variety of clothing ensembles.

For the evening hours, clothing seems similar for all dwellings with the 'long sleeved sweat shirt' being the most frequently used garment. If we compare the morning and evening clothing patterns there seems to be no significant difference. In the evening, there is a complete absence of t-shirt, but still sleeveless t-shirt (which provides even lower thermal protection) is present in A/B and F labelled dwellings. More data are needed in an extended measurement campaign in order to establish detailed clothing patterns of occupants based on the time of the day, their age, sex and health conditions.

§ 5.4.5 Metabolic activity

Figures 5.10 and 5.11 display the metabolic activity levels for the morning and evening hours used for the GSP simulation.

The metabolic activity data during the morning hours show that for the A/B dwellings the most common activity level is 'sitting relaxed' followed by 'lying/sleeping'. For the F labeled dwellings, the most common activity was 'walking' followed by 'light desk work'. Despite the small number of data, which does not allow definite conclusions, the increased metabolic activity (just as with the increased clothing levels), which results in more comfortable thermal sensations, could be another evidence of adjustment for the occupants of the F dwellings.

For the evening hours, the most common metabolic activity of the occupants of A/B labelled dwellings was 'sitting relaxed', while for the F labelled dwellings it was 'walking'. Just like for the morning hours this could be a sign of adjustment to the thermal sensation for the F labelled dwellings' occupants. Two of the three dwellings that recorded 'cool' for thermal sensation had also recorded 'walking' as a metabolic activity despite the fact that indoor temperatures were almost identical for all dwellings. However, the metabolic activities could be related more to the established routines of occupants in the dwellings rather than thermal sensation and further research with increased amount of recorded data is needed.

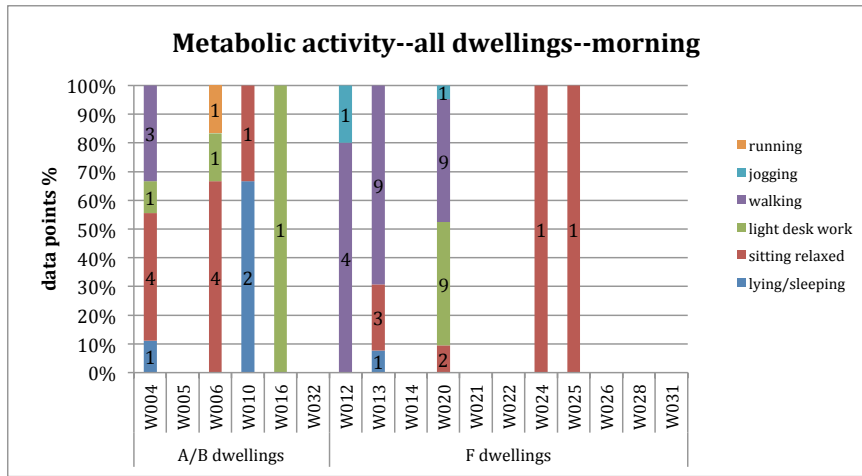


FIGURE 5.10 Morning metabolic activity scores of all dwellings for the total data points used in GSP analysis

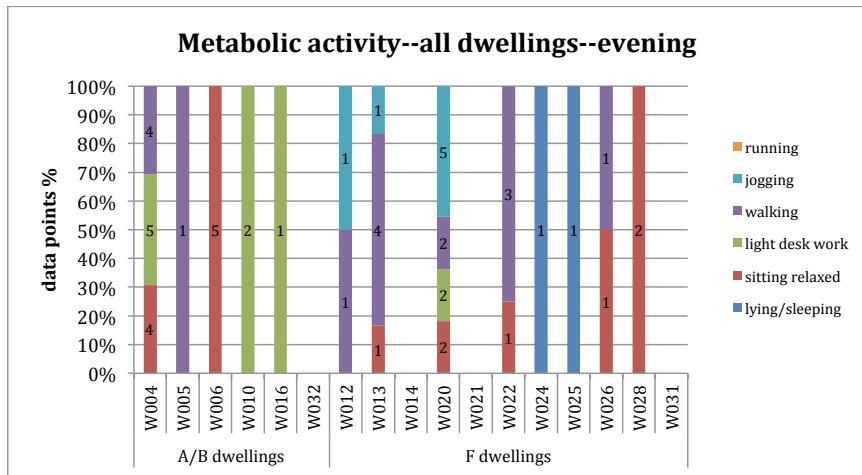


FIGURE 5.11 Evening metabolic activity scores of all dwellings for the total data points used in GSP analysis

§ 5.4.6 Generalized sequential pattern recognition (GSP)

The analysis of the data so far gave us an insight in the cumulative data scores on thermal sensation, indoor temperatures, actions towards thermal comfort, clothing and metabolic activity. However, this analysis is not dynamic, it does not take into account, for example, the exact hour at which an action took place, and what other action, temperature, clothing, and metabolic activity or a combination of the above was recorded at the same hour. Such time combinations between the above-mentioned parameters could also shed light in the causality of certain actions, clothing preferences or metabolic activity patterns. For example if actually metabolic activity is used as an adjustment factor for lower thermal sensations or if warmer clothing is actually used as an adjustment for low temperatures, or if having a warm shower and a hot drink is not related to any of those things and are happening out of pure habit. Moreover, the GSP analysis could lead to patterns supported by all dwellings, which means that with the accumulation of enough data, patterns supported by greater population groups would be possible to be defined.

The data set described in Table 5.2 was fed to the GSP algorithm with the purpose of defining significant sequential patterns. The software that was used for the analysis was rapidminer [22]. The GSP analysis took place for the morning hours between 7-9 a.m. and the evening hours between 5-7 p.m. for all dwellings and for A/B and F label dwellings separately. There is one input string per dwelling per day per timestamp, but the sequences are aggregated on the three morning hours and the three evening hours.

§ 5.4.6.1 Most important sequences

The results of the GSP algorithm concerning the most important sequences discovered for the morning and evening hours are presented in Tables 5.3, 5.4, and 5.5. The events' combinations with the highest support and the smaller amount of events are presented first in the tables. There were many combinations of events that were supported by all dwelling days (Table 5.3), A/B dwelling days (Table 5.4), and F labeled dwelling days (Table 5.5), especially in lower support values such as thirty or twenty per cent. In this study we choose to present results that were supported by minimum of 40% of the dwelling days. In this work, 100% support means that the sequence is found in all dwelling days (meaning in turn that for all days of all dwellings this specific sequence was found between 7 and 9 o'clock).

The sequences (combination of events) are presented as a, b, c etc. meaning that, a was the first event, followed in time by b (although b could also takes place at the same hour as a), followed in time by c (although c could also takes place at the same hour as b).

When all seventeen dwellings were participating in the GSP simulation, for the morning hours, the highest support was found to be 0.59 and the events combination was $20 < T < 22, T > 22$. This means that 59% of the dwelling days between 7-9 a.m. have their temperature increased from a value between 20 °C and 22 °C to a temperature above 22 °C. This combination of events is also the most supported (82%) among the F labeled dwellings. For the evening hours, and for all dwellings participating in the simulation, the most supported sequence (65%) was $T > 22$, Neutral. The same sequence is supported the most by A/B dwellings (67%) and F dwellings (65%). This shows that regardless of the energy label of the dwelling, during the early evening hours, residential dwellers in our sample seem to agree that neutrality is accompanied by temperatures above 22 °C. F label dwellings, however, should consume considerably more energy to reach the same level of indoor comfort.

TABLE 5.3 GSP results from the morning and evening simulation of all dwellings

SUPPORT	EVENTS COMBINATION--MORNING	SUPPORT	EVENTS COMBINATION--EVENING
0.59	20<T<22, T>22	0.65	T>22, Neutral
0.53	20<T<22, A bit cool	0.47	20<T<22, T>22
0.53	T>22, hot drink	0.47	T>22, hot drink
0.53	T>22, warm shower	0.47	T>22, rather warm clothing
0.47	18<T<20, 20<T<22	0.41	20<T<22, Neutral
0.47	20<T<22, thermostat up	0.41	T>22, cold drink
0.47	T>22, A bit cool	0.41	T>22, thermostat up
0.47	T>22, thermostat up	0.41	T>22, sitting relaxed
		0.41	T>22, walking
0.47	20<T<22, T>22, thermostat up	0.41	Neutral, rather warm clothing
		0.41	hot drink, cold drink
0.41	18<T<20, T>22	0.41	T>22, Neutral, rather warm clothing
0.41	20<T<22, Neutral	0.41	T>22, hot drink, cold drink
0.41	20<T<22, hot drink		
0.41	20<T<22, warm shower		
0.41	T>22, Neutral	0.35	T>22, A bit cool
0.41	T>22, rather warm clothing	0.35	A bit cool, Neutral
0.41	A bit cool, Neutral	0.35	Neutral cold drink
0.41	A bit cool, warm shower	0.35	Neutral, sitting relaxed
0.41	hot drink, thermostat up	0.35	rather warm clothing, sitting relaxed
0.41	18<T<20, 20<T<22, T>22	0.35	20<T<22, T>22, Neutral
0.41	20<T<22, T>22, A bit cool	0.35	T>22, A bit cool, Neutral
0.41	20<T<22, T>22, hot drink	0.35	T>22, Neutral, cold drink
0.41	20<T<22, A bit cool, Neutral	0.35	T>22, Neutral, sitting relaxed
0.41	20<T<22, hot drink, thermostat up	0.35	T>22, rather warm clothing, sitting relaxed
0.41	T>22, hot drink, thermostat up		

TABLE 5.4 GSP results from morning and evening simulation of A/B labeled dwellings

SUPPORT	EVENTS COMBINATION--MORNING	SUPPORT	EVENTS COMBINATION--EVENING
0.5	T>22, light desk work	0.67	T>22, Neutral
0.5	A bit cool, A bit warm	0.67	T>22, hot drink
0.5	A bit warm, normal clothing		
0.5	A bit warm, sitting relaxed	0.5	T>22, cold drink
0.5	normal clothing, sitting relaxed	0.5	Neutral, cold drink
		0.5	hot drink, cold drink
0.5	A bit warm, normal clothing, sitting relaxed	0.5	hot drink, normal clothing

>>>

TABLE 5.4 GSP results from morning and evening simulation of A/B labeled dwellings

SUPPORT	EVENTS COMBINATION--MORNING	SUPPORT	EVENTS COMBINATION--EVENING
		0.5	rather warm clothing, light desk work
		0.5	T>22, Neutral, cold drink
		0.5	T>22, hot drink, cold drink

TABLE 5.5 GSP results from morning and evening simulation of F labeled dwellings

SUPPORT	EVENTS COMBINATION--MORNING	SUPPORT	EVENTS COMBINATION--EVENING
0.82	20<T<22, T>22	0.64	20<T<22, T>22
		0.64	T>22, Neutral
0.73	18<T<20, 20<T<22	0.55	T>22, rather warm clothing
		0.55	T>22, thermostat up
0.64	18<T<20, T>22		
0.64	20<T<22, A bit cool	0.45	20<T<22, Neutral
0.64	20<T<22, thermostat up	0.45	T>22, A bit cool
0.64	T>22, hot drink	0.45	T>22, hot drink
0.64	T>22, thermostat up	0.45	T>22, sitting relaxed
0.64	T>22, warm shower	0.45	T>22, walking
		0.45	A bit cool, Neutral
0.64	18<T<20, 20<T<22, T>22	0.45	Neutral, rather warm clothing
0.64	20<T<22, T>22, thermostat up	0.45	rather warm clothing, sitting relaxed
		0.45	20<T<22, T>22, Neutral
0.55	18<T<20, thermostat up	0.45	T>22, A bit cool, Neutral
0.55	20<T<22, Neutral	0.45	T>22, Neutral, rather warm clothing
0.55	20<T<22, hot drink	0.45	T>22, rather warm clothing, sitting relaxed
0.55	20<T<22, warm shower		
0.55	T>22, A bit cool		
0.55	A bit cool, Neutral		
0.55	hot drink, thermostat up		
0.55	18<T<20, 20<T<22, thermostat up		
0.55	18<T<20, T>22, thermostat up		
0.55	20<T<22, T>22, A bit cool		
0.55	20<T<22, T>22, hot drink		
0.55	20<T<22, T>22, warm shower		
0.55	20<T<22, A bit cool, Neutral		
0.55	20<T<22, hot drink, thermostat up		
0.55	T>22, hot drink, thermostat up		
0.55	18<T<20, 20<T<22, T>22, thermostat up		
0.55	20<T<22, T>22, hot drink, thermostat up		

Clearly, there are much more variations (events combinations) in F labeled dwellings than in A/B ones. This could however, result from the significantly higher number of data points related to the F label dwellings.

§ 5.4.6.2 Occupancy Behavior patterns

Such pattern recognition of important sequential events in buildings aims at shedding light in occupancy behavior, related to thermal comfort, which in turn is connected with energy consumption. Having this in mind, we categorized the above combinations of events in two groups that are related to energy consumption, energy and non-energy consuming events, for the morning and evening hours, Table 5.6. Furthermore, the two main categories were further categorized into thermal sensation related and surprising events, which are denoted by superscripts as shown in Table 5.6. By 'energy consuming', we mean all the events that could relate directly to an increase in energy consumption. 'Non energy consuming events' are the events that are not related to an increase in energy consumption. For example the event ($18 < T < 20$, $20 < T < 22$) shows an increase in temperature, which is expected to lead to an increase in energy consumption. Another example are the thermal sensation related events ($20 < T < 22$, Neutral) and ($T > 22$, Neutral). It is logical to expect (despite the numerous parameters that affect thermal comfort) that for temperatures above $20\text{ }^{\circ}\text{C}$ people would have many chances to feel neutral. 'Surprising' were the events that were counter intuitive, having in mind that people would try to maximize their thermal comfort even at the expense of increased energy consumption. For example the events ($20 < T < 22$, $T > 22$), ($20 < T < 22$, thermostat up) or ($T > 22$, A bit cool) describe combinations that are counter intuitive, especially when temperatures are above $22\text{ }^{\circ}\text{C}$ and occupants say they are 'a bit cool' or they turn their thermostat up. Such combinations have more chances to lead to rebound effects and unnecessary energy consumption.

TABLE 5.6 Categorization of combination events in groups related to energy consumption for the morning and evening hours of all dwellings

SUPPORT	MORNING		SUPPORT	EVENING	
	Energy consuming events	Non energy consuming events		Energy consuming events	Non energy consuming events
0.59	20<T<22, T>22 ^{SE}		0.65		T>22, Neutral ^{TS}
0.53		20<T<22, A bit cool ^{TS}	0.47	20<T<22, T>22	
0.53		T>22, hot drink	0.47		T>22, hot drink ^{SE}
0.53	T>22, warm shower ^{TS}		0.47		T>22, rather warm
0.47	18<T<20, 20<T<22		0.41		20<T<22, Neutral ^{TS}
0.47	20<T<22, thermostat up ^{SE}		0.41		T>22, cold drink
0.47		T>22, A bit cool ^{TS,SE}	0.41	T>22, thermostat up ^{SE}	
0.47	T>22, thermostat up ^{SE}		0.41		T>22, sitting relaxed
0.47	20<T<22, T>22, thermostat up ^{SE}		0.41		T>22, walking
0.41	18<T<20, T>22		0.41		Neutral, rather warm ^{TS}
0.41		20<T<22, Neutral ^{TS}	0.41		hot drink, cold drink
0.41		20<T<22, hot drink ^{SE}	0.41		T>22, Neutral, rather warm ^{TS}
0.41	20<T<22, warm shower ^{SE}		0.41		T>22, hot drink, cold drink ^{SE}
0.41		T>22, Neutral ^{TS}	0.35		T>22, a bit cool ^{TS}
0.41		T>22, rather warm	0.35		A bit cool, neutral ^{TS}
0.41	A bit cool, warm shower ^{TS}		0.35		Neutral, sitting relaxed ^{TS}
0.41	hot drink, thermostat up		0.35		rather warm, sitting relaxed
0.41	18<T<20, 20<T<22, T>22		0.35	20<T<22, T>22, Neutral ^{TS}	
0.41	20<T<22, T>22, A bit cool ^{TS,SE}		0.35		T>22, A bit cool, Neutral ^{TS,SE}
0.41	20<T<22, T>22, hot drink ^{SE}		0.35		T>22, Neutral, cold drink ^{TS}
0.41		20<T<22, A bit cool, Neutral ^{TS}	0.35		T>22, Neutral, sitting relaxed ^{TS}
0.41	20<T<22, hot drink, thermostat up ^{SE}		0.35		T>22, rather warm, sitting relaxed
0.41	T>22, hot drink, thermostat up ^{SE}				

TS: thermal sensation related event / SE: surprising event

The most populous category was the 'energy consuming events' with 15 event combinations, followed by 'Surprising events' with 13 event combinations. Even more discouraging, in terms of energy efficiency, is the fact that the energy consuming and surprising events share 10 common events. These unexpected events are mostly related to jumping from already high indoor temperatures to even higher ones. These events are tightly connected with energy consumption and their effectiveness towards thermal comfort is doubtful, given the already very high indoor temperatures. Furthermore, there is a complete absence of alternative ways to improve one's thermal comfort such as clothing, or increased metabolic activity. The GSP algorithm found only one sequence (supported by 41% of the dwelling days nonetheless) for which people feeling 'a bit cool' took a 'warm shower'. However, this is more likely related to a habitual event, since many people have a warm shower in the morning in order to start their day. The combinations of events towards the improvement of thermal comfort showed a prevalence of conventional means such as increase of indoor temperature and turning the thermostat up while actions such as hot drink or warm shower were deemed more as habits rather than actions towards comfort. We have to mention again that the data we had were not exhaustive and that there is a great room for improvement, especially for the gathering of the subjective data such as actions, clothing and metabolic activity.

The GSP simulation for the evening hours showed rather different results compared to the morning hours. The energy consuming combinations were significantly reduced mainly because of the absence of temperatures below 20 °C and having a warm shower. Usually dwellings are not heated during the night and temperatures could fall below 20 °C and even below 18 °C, therefore, it would not be surprising that occupants are trying to increase indoor temperature in the morning hours. Having a warm shower on the other hand seems to be a daily routine more than an action towards comfort. This finding is supported by the results of the chi² tests that are shown in Table 5.4 of chapter 4, according to which for both A/B and F label dwellings, having a "warm shower" was found entirely unrelated to the reported thermal sensation. The 'energy consuming' combinations were reduced to 3 while the 'surprising events' were only 5 and only one of them was shared with the 'energy consuming' category.

§ 5.4.7 Energy+ simulation results

First, the concept house was simulated with the commonly available occupancy profiles and set point temperatures that are predefined in almost every building simulation software such as Energy+, Design Builder, and ESPr. Therefore, the temperature

heating set point was 20 °C for all rooms, and the heating system's availability was matching the occupancy schedule; the heating system was on from 7-9 a.m. when people were waking up and getting ready to go to work. Then it was off until 17:00 when people were absent from the dwelling and on again from 17:00 until 24:00 when people were going to sleep.

Subsequently, the concept house was simulated with the actual hourly temperature profiles and occupancy schedules that we obtained from the measurement campaign. Ioannou and Itard (2015) showed with a Monte Carlo sensitivity analysis, with the same Concept House as the reference building, that using the thermostat and altering the indoor temperature, can explain more than 90% of the variance in the total heating consumption of the dwelling. Therefore, actual hourly heating profiles could improve simulation accuracy compared to business as usual simulations that are taking place with schedules and heating points based on assumptions that may not reflect actual ones

This was done by using the hourly heating profiles of three different types of dwellings that participated in the campaign in order to model a reference dwelling. The dwellings used were A and B label, with gas boiler and radiators as the heating system, A label and heat pump coupled with hydronic underfloor heating, and F label with gas boiler and radiators. As already mentioned in section 3.2 the simulations were repeated three times, one time with the control of the heating system corresponding to the indoor air temperature (T_{air}), one time corresponding to the indoor operative temperature (T_{oper}), and one corresponding to the PMV thermal comfort index. The reason for performing the simulations with the above three different set points was to compare the energy consumption, the indoor temperatures, and the comfort index between these configurations. This approach allows the comparison of the performances of these three control strategies of the heating system.

Because the control set points were not known from the measurement campaign, and only the indoor air temperature was known, the following model calibration procedure was applied:

The actual hourly air temperature profiles from the measurement campaign were fed to the model and the control set points (T_{air} and T_{oper}) were iteratively adjusted up to the moment where the hourly air temperature profiles, resulting from the simulations, were matching the actual ones (the ones obtained during the measurement campaign). When the PMV was used as the control, it was set between -0.5 and +0.5, which corresponds to the neutral comfort level of the PMV scale and the resulting hourly air temperature profile from the simulation is presented in the results and compared to the profiles obtained for T_{air} and T_{oper} as the control set points. The

simulations took place for the period between 1st March and 7th March which is the period that the tenants were handed the comfort dial.

For the reference simulation (standard profile) the T_{air} and T_{oper} were assumed to be 20 °C, during the hours that the dwelling was occupied, which is a common approach among engineers when simulating residential dwellings.

§ 5.4.7.1 A/B label dwellings with boiler and radiators

Figure 5.12 shows the annual heating consumption of the concept house, simulated as an A label dwelling with gas boiler and radiator, with first business as usual schedules and heating set points, and then simulated with the actual hourly heating profiles and occupancy schedules of dwellings W010 and W032. These two dwellings were chosen because they were both in the A/B label category and their actual hourly temperature profiles were above 22 °C and around 20 °C respectively. Figure 5.13 shows the indoor temperature T_{air} and the PMV resulting from the simulations for the living room of those dwellings.

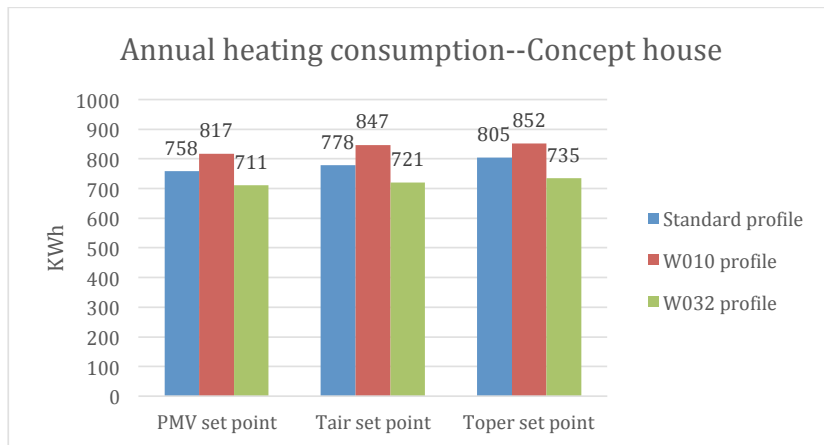
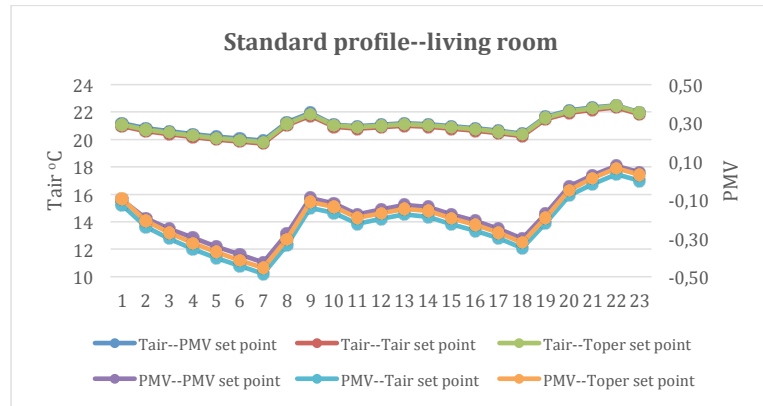


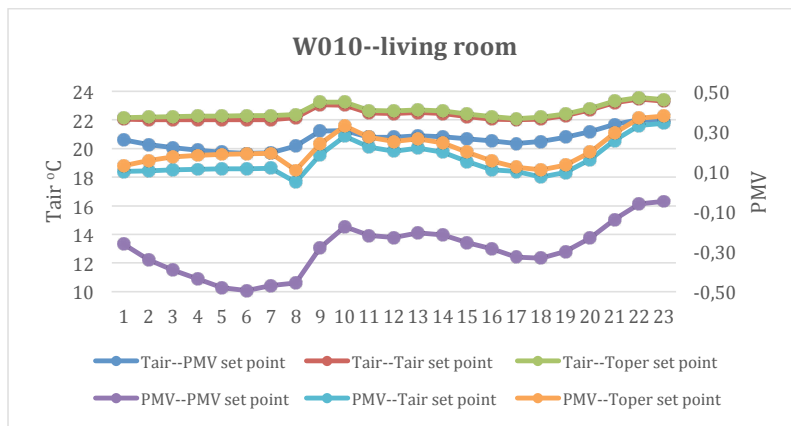
FIGURE 5.12 Annual heating consumption simulated for three different heating set points for the Concept House and dwellings W010 and W032

When heating set point corresponds to the T_{air} (which is the way the majority of thermostats are controlled) or T_{oper} , all profiles lead to higher energy consumption. This clearly relates to the indoor temperatures, Figure 5.13. W010 has the highest indoor

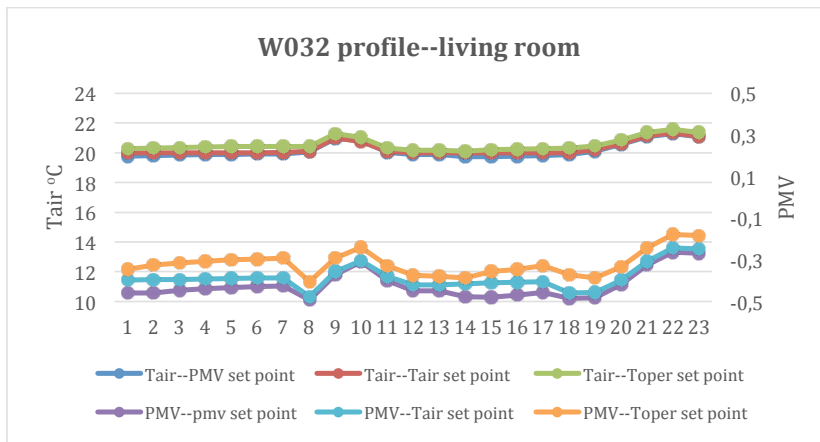
temperature profile, the highest energy consumption, and the most comfortable PMV index, which suggests that the tenants of W010 strive for higher comfort in the expense of energy consumption. However, if the indoor temperature is controlled by the PMV we see that the simulated PMV of tenants is significantly lower (but still within the comfort range) and the indoor air temperature is 1.5 °C to 2 °C lower. This could lead to significant energy savings. This effect, in the presented dwellings, seems to be more obvious when the indoor temperatures of the dwelling are higher. This can be seen in the comparison between W010 and W032. W010 that has the highest indoor temperatures records the greatest drop in the PMV level (and indoor air temperature) when control is switched from T_{air} and T_{oper} to PMV. This effect is smaller (but still significant) in W032.



a



b



c

FIGURE 5.13 Indoor Tair and PMV simulated for the Concept house (well insulated and HR boiler) with three different heating set points and occupancy profiles

§ 5.4.7.2 A label dwellings with heat pump and underfloor hydronic heating

Figure 5.14 shows the annual heating consumption of the concept house, simulated as A label dwelling with heat pump and hydronic underfloor heating system, with business as usual schedules and heating set points, and with the actual hourly heating profiles and occupancy schedules of dwellings W003 and W004. Figure 5.15 shows the indoor T_{air} and PMV for the living room of those dwellings.

The effect of the different heating set points is not visible in this case of dwellings due to the continuous operation of this heating system and the big amount of time needed for specific changes in the thermostat to be felt in the indoor environment of the dwelling. The differences in the annual energy consumption between the dwellings is because of the different hourly temperature profiles that we obtained during the measurement campaign. In the standard profile the concept house was simulated with 20 °C heating set point for the whole day, while W003 and W004 had an average of 26 °C and 24 °C in the living room respectively. The PMV for all dwellings was within the comfort limits and only for concept house, which had the lowest heating set point, the PMV drops slightly below the comfort limits during evening hours. This is due to the undersized heating element that was used for the simulation of each thermal zone of the dwellings (3000 Watts).

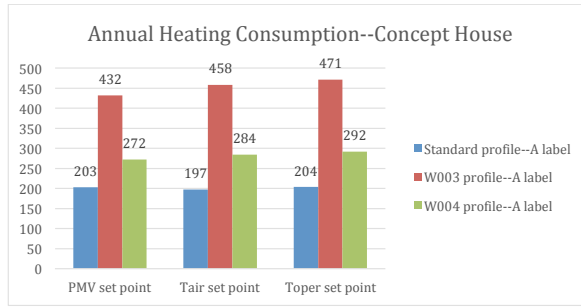
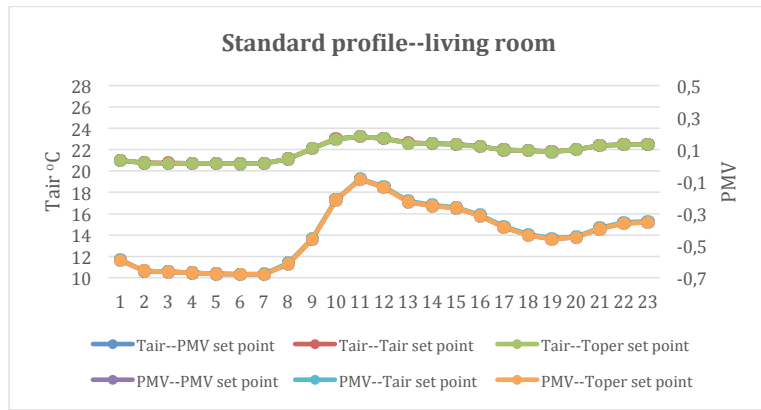
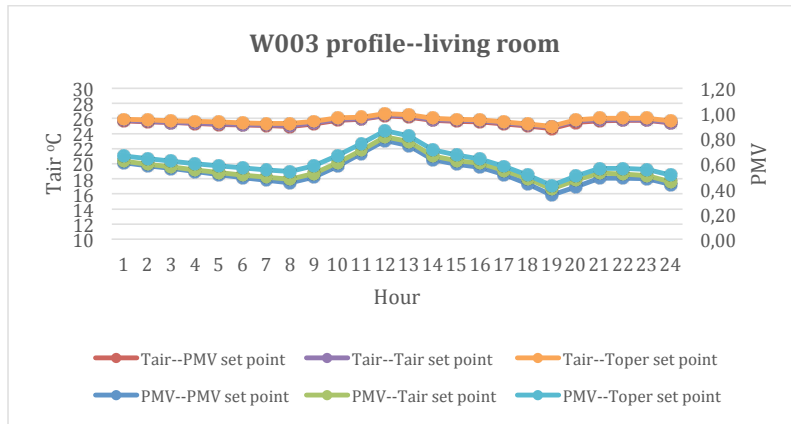


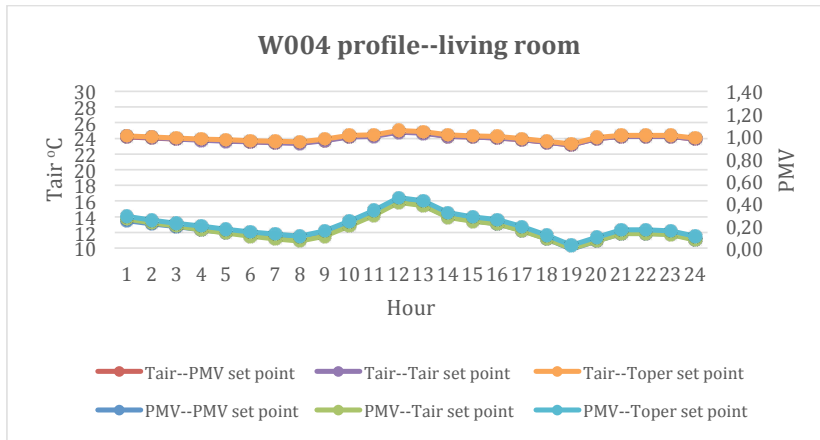
FIGURE 5.14 Annual heating consumption simulated for three different heating set points for the Concept House and dwellings W003 and W004



a



b



c

FIGURE 5.15 Indoor Tair and PMV simulated for the Concept house (well insulated, heat pump, and underfloor heating) with three different heating set points and occupancy profiles

§ 5.4.7.3 F label dwellings with boiler and radiators

Figure 5.16 shows the annual heating consumption of the concept house, simulated as an F label dwelling, with gas boiler and radiator, with business as usual schedules and heating set points, and simulated with the actual hourly heating profiles and occupancy schedules of dwellings W013, W022, W026, and W031. Figure 5.17 shows the indoor T_{air} and PMV for the living room of those dwellings.

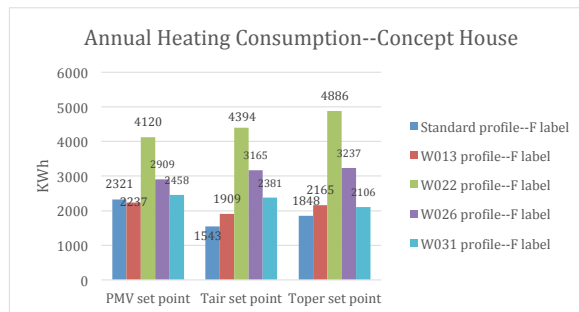
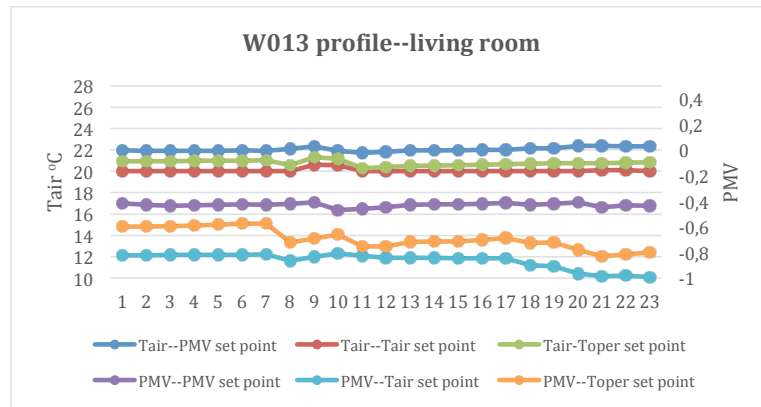
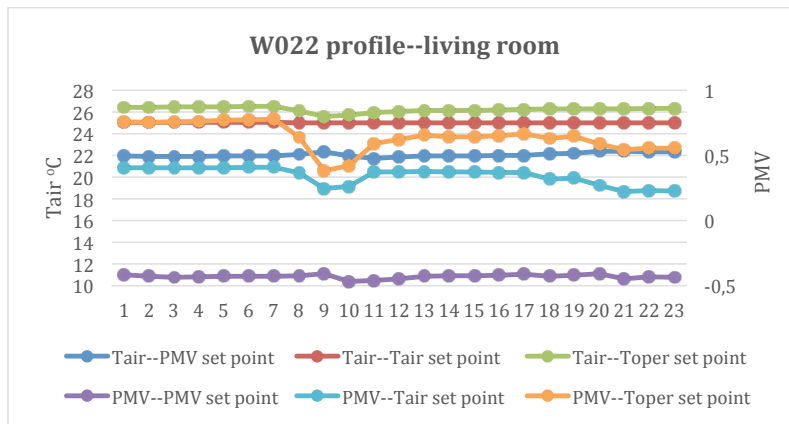


FIGURE 5.16 Annual heating consumption simulated for three different heating set points for the Concept House and dwellings W013, W022, W026, and W031

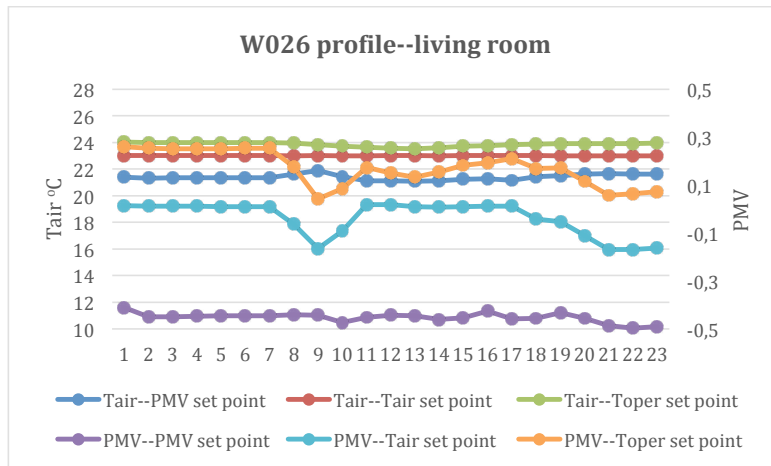
Using the PMV set point as the corresponding value for the operation of the heating system results in the lower energy consumption in W022 and W026. The reason for this is, similar to the case of A label dwellings (Figures 5.12 and 5.13), the unusual high temperature profiles preferred by the tenants of these dwellings, Figure 5.17. As we can see in the graph for dwelling W022 the indoor air temperatures are above 24 °C for the whole day, while for maintaining an hourly comfort level of -0.5, only 22 °C are needed, Figure 5.17. In contrast, W013 has lower indoor temperatures for the whole day and the PMV calculations show that tenants are not supposed to be feeling neutral. In this case, switching to PMV as the set point will result to increased energy consumption, which, however, will bring the tenants within the comfort zone of the PMV index. Nonetheless, during the evening hours the tenants of W013 reported neutral thermal sensations just like their W022 counterparts. This suggests that they might have adjusted their thermal comfort levels to a lower level compared to the tenants of W022 or that the later are more comfortable than they need, utilizing a rebound effect on comfort. Therefore, using the PMV as the set point temperature could result to either an increase or decrease in the energy consumption, depending in the indoor temperature that the tenants prefer. In any case, the comfort of the tenants in this case will be brought within the comfort zone of the Fanger model. But as we saw for the example of W013, this could not be the desired comfort level of the tenants.



a



b



c

FIGURE 5.17 Indoor Tair and PMV simulated for three different heating set points for the Concept House and dwellings W003 and W004

Majcen et al. [27] demonstrated the discrepancy between actual and calculated energy consumption in energy labelled residential dwellings in the Netherlands. Furthermore, Santin [33] and Page et al. [2] showed the importance that occupancy behavior might have in the energy consumption of a dwelling. From a building simulation perspective, Ioannou and Itard [23] showed that behavioral parameters such as the use of the thermostat affects greatly the total energy consumption and the PMV of the tenants. Therefore, if the tenants of a residential dwelling command their indoor environment based on their comfort levels, the components of building simulation software related to the PMV must be improved.

In order to calculate the PMV index, values from six parameters are needed; clothing, metabolic activity, mean radiant temperature, air speed, air temperature, and relative humidity. In a smart built environment, it would be easy to gather the quantitative data related to the PMV with the use of an extensive network of sensors. However, clothing and metabolic activity are more difficult to capture, but a mobile or tablet application incorporating the features of the comfort dial and log book, could give a solution to this problem. Gathering enough subjective data and simulating them with the GSP algorithm could lead to hourly clothing and metabolic activity profiles that would improve greatly the simulation components related to the PMV, thus, improving the accuracy of the simulated energy consumption of residential dwellings.

§ 5.5 Conclusions

Using big data, from a sensor rich environment in residential dwellings, into a data driven model such as the GSP algorithm could lead to the prediction of occupancy behavior patterns. Even grouping all dwellings together, regardless of the energy label, provided high enough support (% of dwelling days that are following a pattern in a specific hour) for occupancy patterns that were revealed by the simulation. For example, in 59% of dwelling days in the morning hours the temperatures between 7-9 a.m. were increasing from $20^{\circ}\text{C} < T < 22^{\circ}\text{C}$ to $T > 22^{\circ}\text{C}$. Furthermore, in 56% of them the temperature $20^{\circ}\text{C} < T < 22^{\circ}\text{C}$ was found to be a bit cool and even for temperatures above 22°C occupants were reporting having a warm shower leading to the suspicion that a warm shower is a routine action not related to thermal comfort. For the evening hours between 5-7 p.m. the simulation for all dwellings showed that in 65% of the dwelling days temperatures higher than 22°C were found to be neutral and in half of them the temperature was increased from $20^{\circ}\text{C} < T < 22^{\circ}\text{C}$ to $T > 22^{\circ}\text{C}$. For only the A/B label dwellings, GSP showed that in 80% of the dwelling days temperatures above 22°C were experienced as being neutral. Furthermore, in the F labeled dwellings in 64% of the dwelling days $T > 22^{\circ}\text{C}$ was found to be neutral and the temperature was increased from $20^{\circ}\text{C} < T < 22^{\circ}\text{C}$ to $T > 22^{\circ}\text{C}$. This shows that tenants of lower labeled dwellings do not compromise their comfort by heating less than the tenants of A/B label dwellings. This will lead of course to higher energy consumption. This is in agreement with some of the findings of the initial questionnaire given to the tenants. To the question "do you find it difficult to pay you monthly energy bills?" all tenants replied "no" despite the fact that the household incomes ranged between 700 to 4.5 thousand euros.

Furthermore, the sequential pattern analysis revealed patterns of occupancy behavior that were categorized as energy consuming, non-energy consuming, thermal sensation related, and surprising. The common notion in building simulations, reflected in the premade models of occupancy available in simulation software, is that during the night the heating is switched off, temperature drops and therefore in the morning hours when people wake up they try to bring the temperature to the desired comfort level. However, the hourly air temperature profiles of the specific dwellings mentioned in this study suggest otherwise since the temperature profiles during the night were very stable and most of the time above 20 °C. If the “energy consuming” patterns are due to habitual reasons then a GSP algorithm could reveal these patterns and feed them back to the tenants leading to potential energy savings, as long as of course these patterns do not compromise their comfort levels.

Finally, the GSP pattern recognition could be proven beneficial in the improvement of the building simulation process. Subjective parameters that are very difficult to capture and transform into hourly profiles, to be used in simulations, can be fed to the GSP algorithm, via information technology applications for mobile phones or tablets, and can be processed into hourly profiles. These customized profiles can afterwards be used to predict more accurately the energy consumption of a specific dwelling. If common patterns are found between large groups of dwellings then profiles that are more generic can be created for larger groups of dwellings based on their energy label, heating system or other categories.

Propositions for further research include the development of a more detailed application for smartphones or tablets for the tenants. The more data are fed into the algorithm, the more its precision will improve and therefore a more exhaustive, non-obligatory, selection of choices should be available. Furthermore, a challenging task would be how the findings of the GSP algorithm could be used. Some people might be interested in reducing their energy consumption while others might be interested in maximizing their comfort, or some might be interested in finding a balance between the two. The findings of the GSP could be used to attempt to alter tenants' behavior by introducing a teaser function in order to save energy, or they could just be used for tenants to help them find the appropriate levels of indoor parameters to maximize their comfort. Moreover, the customized profiles obtained by the GSP algorithm should be used in an attempt to close the gap between the simulated and actual heating consumption in residential dwellings.

References

- REF. 5.01 Heierman, Ed, M. Youngblood, and Diane J. Cook. "Mining temporal sequences to discover interesting patterns." *KDD Workshop on mining temporal and sequential data*. 2004.
- REF. 5.02 Page, Jessen, et al. "A generalized stochastic model for the simulation of occupant presence." *Energy and buildings* 40.2 (2008): 83-98.
- REF. 5.03 Dong, Bing, and Burton Andrews. "Sensor-based occupancy behavioral pattern recognition for energy and comfort management in intelligent buildings." *Proceedings of building simulation*. 2009.
- REF. 5.04 Fritsch, R., et al. "A stochastic model of user behavior regarding ventilation." *Building and Environment* 25.2 (1990): 173-181.
- REF. 5.05 Degelman, Larry O. "A model for simulation of day lighting and occupancy sensors as an energy control strategy for office buildings." *Proceedings of building simulation*. Vol. 99. 1999.
- REF. 5.06 Reinhart, Christoph F. "Lightswitch-2002: a model for manual and automated control of electric lighting and blinds." *Solar Energy* 77.1 (2004): 15-28.
- REF. 5.07 Wang, Danni, Clifford C. Federspiel, and Francis Rubinstein. "Modeling occupancy in single person offices." *Energy and buildings* 37.2 (2005): 121-126.
- REF. 5.08 Mahdavi, Ardeshir, et al. "User interactions with environmental control systems in buildings." *Proceedings PLEA*. 2006.
- REF. 5.09 Lam, Khee Poh, et al. "Occupancy detection through an extensive environmental sensor network in an open-plan office building." *IBPSA Building Simulation* 145 (2009): 1452-1459.
- REF. 5.10 Cook, Diane, and Sajal Kumar Das. *Smart environments: Technology, protocols and applications*. Vol. 43. John Wiley & Sons, 2004.
- REF. 5.11 <http://monica.ir.nl/>.
- REF. 5.12 www.SusLabNWE.eu.
- REF. 5.13 <http://installaties2020.weebly.com/>.
- REF. 5.14 Ioannou, Anastasios, and Laure Itard. "In-situ and real time measurements of thermal comfort and its determinants in thirty residential dwellings in the Netherlands." *Energy and Buildings* 139 (2017): 487-505.
- REF. 5.15 In-situ real time measurements of thermal comfort and comparison with the adaptive comfort theory in Dutch residential dwellings.
- REF. 5.16 Youngblood, G. Michael, and Diane J. Cook. "Data mining for hierarchical model creation." *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 37.4 (2007): 561-572.
- REF. 5.17 Gupta, Manish, and Jiawei Han. "Applications of pattern discovery using sequential data mining." *Data Mining: Concepts, Methodologies, Tools, and Applications*. IGI Global, 2013. 947-969.
- REF. 5.18 Agrawal, Rakesh, and Ramakrishnan Srikant. "Mining sequential patterns." *Data Engineering, 1995. Proceedings of the Eleventh International Conference on*. IEEE, 1995.
- REF. 5.19 H. Mannila, H. Toivonen, and A. Verkamo. Discovering frequent episodes in sequences. In Proc. 1st International Conference on Knowledge Discovery and Data Mining (KDD'95), pp. 210-215, Montreal, Canada, August 1995.
- REF. 5.20 Agrawal, Rakesh, and Ramakrishnan Srikant. "Fast algorithms for mining association rules." *Proc. 20th int. conf. very large data bases, VLDB*. Vol. 1215. 1994.
- REF. 5.21 Srikant, Ramakrishnan, and Rakesh Agrawal. "Mining sequential patterns: Generalizations and performance improvements." *International Conference on Extending Database Technology*. Springer Berlin Heidelberg, 1996.
- REF. 5.22 <https://rapidminer.com/>
- REF. 5.23 Ioannou, A., and L. C. M. Itard. "Energy performance and comfort in residential buildings: Sensitivity for building parameters and occupancy." *Energy and Buildings* 92 (2015): 216-233.
- REF. 5.24 Dhar, Vasant. "Data science and prediction." *Communications of the ACM* 56.12 (2013): 64-73.
- REF. 5.25 <http://www.opschaler.nl/>
- REF. 5.26 Kornaat, Wim, et al. "Development of improved models for the accurate prediction of energy consumption in dwellings." (2016).
- REF. 5.27 Majcen, D., L. C. M. Itard, and H. Visscher. "Theoretical vs. actual energy consumption of labelled dwellings in the Netherlands: Discrepancies and policy implications." *Energy policy* 54 (2013): 125-136.
- REF. 5.28 Dear, R. de, Gail Brager, and Donna Cooper. "Developing an adaptive model of thermal comfort and preference-Final Report (ASHRAE RP-884)." *Atlanta, GA: ASHRAE* (1997).
- REF. 5.29 Bouden, Chiheb, and Nadia Ghrab. "An adaptive thermal comfort model for the Tunisian context: a field study results." *Energy and Buildings* 37.9 (2005): 952-963.

- REF. 5.30 Heidari, Shahin, and Steve Sharples A comparative analysis of short-term and long-term thermal comfort surveys in Iran, *Energy Build.* 34.6 (2002)607–614.
- REF. 5.31 Joseph Khedari, Boonlert Boonsri, Jongjit Hirunlabh, Ventilation impact of a solar chimney on indoor temperature fluctuation and air change in a school building, *Energy Build.* 32.1 (2000) 89–93.
- REF. 5.32 <https://www.cbs.nl/en-gb/news/2007/27/average-income-dutch-household-approximately-50-thousand-euro>
- REF. 5.33 Santin, Olivia Guerra, Laure Itard, and Henk Visscher. “The effect of occupancy and building characteristics on energy use for space and water heating in Dutch residential stock.” *Energy and buildings* 41.11 (2009): 1223-1232.
- REF. 5.34 Santin, Olivia Guerra. “Behavioral patterns and user profiles related to energy consumption for heating.” *Energy and Buildings* 43.10 (2011): 2662-2672.