Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institut für
Technische Informatik und
Kommunikationsnetze

Indoor WSN with On-Board PV Energy Harvesting - Making Sense of People Detection Data

Semester Thesis

Markus Kiser

kiserm@ethz.ch

Computer Engineering and Networks Laboratory
Department of Information Technology and Electrical Engineering
ETH Zürich

Supervisors:
Stefan Drašković
Naomi Stricker
Prof. Dr. Lothar Thiele

January 8, 2021

Acknowledgements

This project gave me the possibility to work on an intriguing topic in a fast emerging technology field. Thanks to the collaboration with the Institute of Computer Engineering and Networks Laboratory at the Swiss Federal Institute of Technology (ETH Zurich), I was granted the experience to conduct this project in the fascinating field of indoor wireless sensor networks (WSN) with on-board photo-voltaic (PV) energy harvesting.

My special gratitude goes to my supervisors Stefan Drašković and Naomi Stricker. I highly appreciate their valuable support, their knowledge and their uncountable inputs. I especially would like to thank them for their great flexibility during the challenging situation with COVID-19. Without that, it would not have been possible to complete this project the way I did.

Zurich, January 8, 2021

Markus Kiser

Abstract

Wireless Sensor Networks (WSN) are currently receiving great attention. Due to the increasing number of low power sensor systems, it is important to have solutions how to optimally deploy them on a big scale regarding cost-effectiveness. One approach is to have on-board energy harvesting solutions such that the effort for the deployment as well as the maintenance of a WSN can be drastically reduced. One issue of this approach arises when WSN are deployed indoor, because very little data is available about the indoor harvestable energy. Moreover, the indoor harvested energy is very volatile. This makes resource-efficient system design and operation a hard task.

This work examines, how people presence in a room and seasonal differences influence the amount of the indoor harvested photo-voltaic (PV) energy. All energy measurements were collected with an already existing system. However, no ground truth data regarding room occupancy was available. Hence, three different approximations of the ground truth were analyzed. Firstly, whole years energy measurement data traces were examined with respect to seasonal differences and with respect to weekdays (people present) versus weekends (people absent). Secondly, the corona lockdown period in spring 2020 in Switzerland was compared to the same period of previous years. Hence, the people absence ground truth was very well approximated in this approach. Thirdly, a low power sensor network with passive infrared (PIR) sensors was deployed to not only have a good people absence but also presence approximation of the ground truth.

All the data was analyzed qualitatively by conducting a principal component analysis (PCA) as well as quantitatively by calculating statistical indicators. Additionally, two simple energy prediction schemes were evaluated, of which one considered the people presence data and the other did not.

It is found that in settings, where most of the available light is originated from artificial light sources, people present lead to an increase in the harvested energy. In settings where most of the available light is originated from natural light, people present have a minor or no influence. In those cases seasons (spring, summer, fall or winter) have a stronger impact than people presence on the amount of indoor harvested PV energy.

Contents

List of Acronyms

CNN Convolutional Neural Network.

EWMA Exponentially Weighted Moving Average.

PC Principle Component.

PCA Principle Component Analysis.

PIR Passive Infrared.

PV Photo-Voltaic.

UDP User Datagram Protocol.

UTC Coordinated Universal Time.

WSN Wireless Sensor Network.

Chapter 1

Introduction

In recent years, a lot of wireless sensor networks (WSN) were installed. However, cost-effective installing and maintaining of a high number of low-cost sensor systems is a challenging task so far, because an energy source for these systems is needed. Thus, there are several possibilities how to power them. A first possibility is to connect them statically to the local power grid. Thus, there is no or very little need for maintenance. However, for the initial installation a big effort has to be made. Another option is to deploy them with non-rechargeable batteries such that the initial deployment can be easily done. However, one drawback lies in the high maintenance costs, because the batteries have to be replaced regularly. A third and very interesting possibility is a combination of a non-rechargeable and rechargeable battery with an on-board energy harvesting solution at the sensor node. Thus, low initial installation costs are combined with low maintenance costs due to the fact that the systems most of the time draw the energy from the rechargeable battery and only use energy from the non-rechargeable battery in emergency cases.

System designers need to have some knowledge about the harvestable PV energy in order to be able to select appropriate system components. WSN can be deployed outdoor as well as indoor. There is a lot of historical and meteorological data available for the outdoor solar harvesting applications [1]. Yet, there are hardly any data sets for indoor solar harvested energy available. Moreover, it has not been done to a large extent to determine the factors, which influences the harvested solar energy indoors. Additionally, the harvestable energy is very volatile in indoor environments, which makes resource-efficient system design a hard task. Hence, it is crucial to gain more knowledge about the harvested PV energy in indoor environments to make the use of resources more efficient.

Sigrist et al. have collected indoor PV harvested energy data for over three years at ETH Zurich [2], but this data has not yet been analyzed deeply. Besides this, a previous student work [3] showed in a proof of concept manner that people detection and also people counting could be implemented on a system with two low power nodes with the help of passive infrared (PIR) sensors.

The aim of this project is to determine how people presence in a room and seasonal differences influence the amount of indoor PV harvested energy. Since there is no ground truth data about the people presence and absence available, three different approximations of the ground truth were used in order to get a best possible answer of the before mentioned question. Firstly, whole years data traces were examined with respect to weekdays (people present) versus weekends (people absent) and with respect to seasonal differences. Secondly, the corona lockdown period in spring 2020 in Switzerland was compared to the same period of previous years. Hence, the people absence ground truth was very well approximated in this second approach. Thirdly, a low power sensor network with passive infrared (PIR) sensors was deployed to not only achieve a good people absence but also presence approximation of the ground truth.

The energy measurement data set of Sigrist et al. [2] was analyzed qualitatively by conducting a principal component analysis (PCA) as well as quantitatively by calculating statistical indicators. Finally, two simple energy prediction schemes based on an exponentially weighted moving average (EWMA) approach, of which one considered the people presence data and the other did not, were compared. The corresponding people presence data was collected by the aforementioned low power people detection system.

CONTRIBUTION In this project the aforementioned low power people counting system [3] is deployed at ETH. In addition, various people detection algorithms that only take one PIR sensor as an input are implemented within that system. The details of all the contributed implementations within that system are described in Section 4. Having the system set up and installed, data is collected. Besides this, the indoor PV harvesting data set from ETH [2] is examined in depth. In the final part of this project, the collected people detection data is combined with the existing harvested indoor PV energy data from ETH. Finally, two simple energy prediction schemes based on an EWMA approach, of which one considers the collected people presence data and the other does not, are implemented for different prediction horizons (five minutes, ten minutes and one hour) and are compared to each other.

Chapter 2

Related Work

This chapter gives a brief overview of previous work in the field of low power people detection systems as well as in the field of indoor PV harvesting without claiming to be exhaustive.

Reynaud et al. [4] found out that organic PV cells outperform commercial silicon cells in indoor lighting conditions even though under standard test condition it is the other way around. Moreover, the fabrication process of organic cells is low cost and could be put on flexible substrates which offers further benefits for indoor use. Besides this, they state that placing the solar cell optimally to maximize power production is not straight-forward since placing it directly under the light source is seldom practical.

In another work, Ma et al. [5] examined a power estimation method which takes spectral and intensity information of the light into account to optimize the component choices for indoor light energy harvesting systems. They state that to dimension and design an energy harvesting system, one needs good estimates of the available ambient energy, which can be harvested. With this knowledge, suitable PV cells and energy storage elements can be chosen.

One approach to tackle the people counting problem on a low power system is using thermal images to detect people with the help of a convolutional neural network (CNN), which is small enough to run on a limited-memory low-power platform [6]. However, Pratama et al. [7] mention in their work that in a typical office space PIR sensors are the most common technology to detect occupancy. The sensors can only give binary information if the room is occupied or not, but people can not be identifiable.

Wahl et al. [8] conducted an experiment where they tried to estimate people count per office space by using cleverly placed PIR sensors. According to them one benefit of the PIR sensors is that they are mass-produced at very low cost.

However, one disadvantage is that these sensors have a masking time. During this time the PIR sensors can not detect any further movements. This can lead to errors in time sensitive applications like people counting, if e.g. more than one person passes through the observed gateway. This was also considered to be the most prominent error in their work. Besides this, Wahl et al. suggest that their PIR sensor node prototype is able to robustly identify movement in a real world setting and that this technology has the potential to be used in people counting applications.

In another work, Yoon et al. [9] examined the use of PIR sensors along with a door sensor to determine the accuracy of people occupancy detection in a single room. They state that PIR sensors are one of the simplest and most cost-effective approaches to detect people occupancy in a room. Interestingly, they found out that a PIR sensor located on a wall produces a more precise detection rate in comparison to those located on the ceiling or on the wall besides the door. Moreover, using only one sensor at the right location produced more accurate results for detecting occupancy than using three PIR sensors. Overall they found out that it is difficult to measure occupancy in a room by only using PIR sensors. However, if a door sensor is used in addition, then the accuracy could be improved.

The main contributions of this work is not to find the accuracy of a specific low power PIR sensor system to detect occupancy of a room. Nor is this work focused on specifying and analyzing the harvested PV energy for different indoor lighting conditions. This work focuses much more on the question how people occupancy influence the indoor harvested energy in a room. It investigates, if more indoor PV energy is harvested when people are present compared to when people are absent in a single room. How this problem is tackled, is described in section 3.

Chapter 3

Method

There is no ground truth data about the room occupancy. Hence, the strategy, which was used to find an answer to the question how people occupancy and seasonal differences influence the indoor harvested PV energy, is described and reasoned in this chapter.

Firstly, the ETH data set [2] is analyzed in depth. As Figure 3.1 and Table 3.1 show, the measurement locations vary quite a lot in terms of their mixture of artificial and natural light sources.

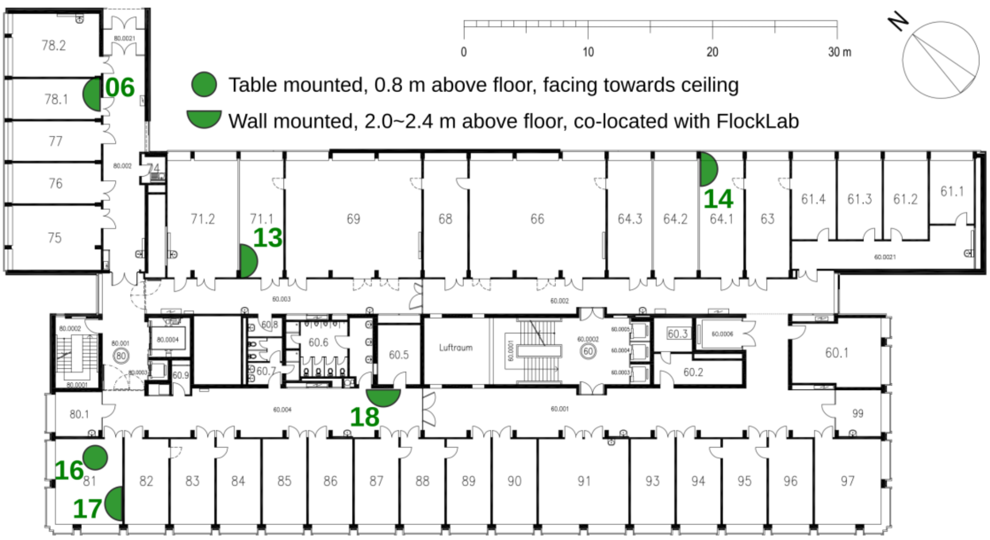


Figure 3.1: Floorplan with the deployed indoor measurement locations. The rounded part of the position markers indicate the directions from which the light reaches the solar panel. [2].

Table 3.1: Characterization of each measurement station from [2]. The names are added and defined to simplify the room characteristics in short names. 'Bright' actually means that the harvested PV energy is mostly driven by natural light whereas 'dark' actually means that the harvested PV energy is mostly driven by artificial light.

Pos.	Description	Roughly Daily Harv. Energy
06	• Employee office • wall mounted at 2.4 m above floor level • little natural light • no direct sun exposure • Name: dark office	2.02 ± 1.64 J
13	• Student office • wall mounted at 2.0 m above floor level • some natural light • no direct sun exposure • Name: dark student room	1.57 ± 1.28 J
14	• Laboratory • wall mounted at 2.1 m above floor level • significant natural light • potential direct sun exposure • Name: bright lab	14.18 ± 11.67 J
16	• Employee office • table mounted and facing towards ceiling • significant natural light • no direct sun exposure • Name: bright office on table	7.07 ± 1.50 J
17	• Employee office • wall mounted at 2.4 m above floor level • significant natural light • no direct sun exposure • Name: bright office on wall	2.87 ± 2.36 J
18	• Hallway • wall mounted at 2.4 m above floor level • no natural light • only little, indirect artificial light • Name: hallway	0.19 ± 0.12 J

Since at natural light driven positions the harvested indoor PV energy might also depend on the outdoor brightness, not only the influence of people presence is examined, but also the effect of the seasonal changes is investigated. By looking at whole years of data, the seasonal differences in the amount of the harvested energy between spring, summer, fall and winter are compared. When analyzing the data with respect to the people presence, there is no ground truth. Thus, for each room the best approximation of the ground truth is to be defined. For all the measurement stations except for station 16 whole years of comparable harvested energy data is available. Table 3.2 lists the considered energy measurement raw data time periods of each measurement station.

Table 3.2: The considered energy measurement raw data time periods of each measurement station are listed. There would have been more data available for the position 18. However, because of a change in lighting on the 22^{th} of October in 2019, the energy measurement data after that day could no longer be compared to days before that day.

Position	Start Date Raw Data	End Date Raw Data
06	2017-10-10	2019-06-17
13	2017-10-10	2020-10-27
14	2017-10-10	2020-10-27
16	2019-06-18	2020-11-29
17	2017-10-10	2020-11-29
18	2017-10-10	2019-10-21

Using whole years of data for each station except of station 16 the difference between weekdays and weekends are examined, because it is assumed that during weekdays people are present and during weekends not. This is for measurement stations 06 and 18 the only available approximation of the ground truth. However, for measurement stations 13, 14 and 17 an even better approximation of the negative ground truth is available thanks to the corona lockdown in spring 2020 in Zurich. During this time it is known with high probability that no people were inside the rooms. However, this high certainty is only for times when no people are present. During the same period of previous years, by which the corona lockdown data is compared, the certainty that the rooms are occupied is assumed to be the same as in the analysis of the whole years during weekdays.

To have an even better approximation of the negative (people absent) and positive (people present) ground truth, a low power people detection system (see section 4) is deployed. A low power system with PIR sensors is intentionally used. A turnstile, light barriers or video surveillance could have been also used with the benefit of a very low detection error. However, if people presence would have been detected with such an approach, which does not fit on a low power sensor node, more work has to be done afterwards to also proof that a people detection system could be implemented within a low power sensor network.

However, if a low power system for occupancy detection is used from the beginning, the before mentioned two steps can be done in a single step. However, this approach has the drawback of having less accuracy in people detection.

The low power system is deployed only at the door of room 81, in which the measurement stations 16 and 17 are located (see Figure 3.1). With that one out of five measurement rooms could be complemented by not only having energy measurement data but also having data which indicates if people are present or absent in this specific room during roughly one month. It is assumed that for artificial light driven positions the differences in harvested energy with respect to people presence or absence can be seen more easily than for natural light driven positions. Consequently, it is assumed to be more interesting in a first investigation of the indoor solar harvested energy data set to have a closer look at locations, which are more driven by natural light. Because the difference between absence and presence of people in such rooms with respect to the harvested energy is assumed to be much more subtle. In details that means for example that if in these rooms only whole years data and corona lockdown data are analyzed, it might happen that no difference in the harvested energy at all can be observed. But with the help of a people detection system it might be possible to see the influence of people present on the harvested energy. Consequently, the *bright office* with measurement stations 16 and 17 as well as the *bright lab* with station 14 inside are candidates to place the people detection system. Finally, the *bright office* is chosen to deploy the system because of two reasons:

- Since in Fall 2020 the corona situation in Zurich was still fragile, the *bright lab* could have been empty for most of the time due to another possible home schooling recommendation by the government. However, the *bright office* would not be affected by that, because it is occupied by some working PhD students.
- In the *bright office* two measurement stations 16 and 17 are located, which made it more interesting to compare two different energy measurements depending on the absence or presence of people in the room compared to the *bright lab*, which has only one measurement station in its room.

In order to predict the future availability of PV energy, the collected people detection data is combined with the existing indoor harvested PV energy measurement data from ETH. Two simple energy prediction schemes are implemented based on an EWMA approach, of which one considers the collected people presence data and the other does not. They are compared to each other in terms of their absolute energy prediction errors. Moreover, both schemes are implemented for three different prediction horizons of five minutes,ten minutes and one hour.

CHAPTER 4

People Detection System

In this chapter a closer look is taken at the people detection system. It is used for having data on room occupancy. Since there is no ground truth data about room occupancy, this system gives a good approximation on the negative (people absent) and positive (people present) ground truth. However, this system will never give perfect ground truth data, because of its limitations (see Section 4.3). This chapter starts with describing the system state at the beginning of this work. In Section 4.2 the extensions of the system, which are implemented during this work, are described. The different binary filters are described in more detail in Section 4.2.1. Lastly, Section 4.3 describes the limitations of this low power detection system with PIR sensors.

4.1 Initial System

The system was build and set up by a previous student work [3]. The technical details of the system can be found in that report. Here only the main parts of it are summarized. The people counting system consists of three main parts:

- Two Permamotes as Wireless Sensor Network (WSN) end nodes
- nRF52840-Dongle as Network Co-Processor
- Raspberry Pi 3 Model B+ as Border Router

In another local environment, this system composed of this three components could already detect motion and count people. If some motion was recognized by the PIR sensor on the Permamote, a packet is send via a wireless connection with the user datagram protocol (UDP) to the nRF52840-Dongle. After receiving the packet, the information is propagated locally inside the border router to do some post-processing with that packet (parsing, analyzing, storing, etc.) Besides this, a first Web-GUI was implemented in Python on the border router by using the dash library.

4.2 Enhancements

The code structure of the initial system from the student project [3] was only consisting of two executable python scripts. One was the *gui_people_detector.py* and the other was the *packet_analyzer.py*. During this work some extensions are made to the post-processing of the received data packets from the Permamotes and by that the Python code is modularized. All sensor data (PIR, light, voltage of the solar cell, primary and secondary battery) of each Permamote is stored locally on the border router in csv files. In addition to that, four different binary filters (see Section 4.2.1) and three different people counting algorithms (see Section 4.2.2) are implemented. The Python project structure on the border router for the post-processing looks as follows (where the scripts marked with a star (*) are executable).

- gui_people_detector.py (*)
- packet_analyzer.py (*)
- test_utils.py (*)
- utils
  - binary_filter.py
  - detection_helper.py
  - gui_helper.py
  - people_counter.py
  - config.py
- data
  - YYYY-MM-DD
    - any_motion_list_YYYY-MM-DD.csv
    - motion_list_id_X_YYYY-MM-DD.csv
    - people_counter_YYYY-MM-DD.csv
    - binary_filter_id_X_specific_YYYY-MM-DD.csv
    - binary_filter_id_X_exp_no_sum_YYYY-MM-DD.csv
    - binary_filter_id_X_exp_simple_YYYY-MM-DD.csv
    - binary_filter_id_X_exp_sum_last_2_YYYY-MM-DD.csv
    - light_lux_list_id_X_YYYY-MM-DD.csv
    - volt_bat_id_X_YYYY-MM-DD.csv
    - volt_sec_id_X_YYYY-MM-DD.csv
    - volt_sol_id_X_YYYY-MM-DD.csv

The script *packet_analyzer.py* has to run constantly on the border router such that the packets from the Permamotes can be received, parsed and processed. It is also recommended to run constantly the script *gui_people_detector.py* such that the GUI can be opened in a browser (see Figure 4.1) at every moment in time and from everywhere, because the border router has access to the internet. Because new data with the binary filters is available, the GUI is extended in a way that not only the people counter data but also the binary data is visualized. Moreover, the GUI is improved such that the user can input the date of which the corresponding data is displayed. An important detail which is worth mentioning is that the timestamps of the received data packets from the Permamotes is changed from UTC time to the local *Europe/Zurich* timezone.

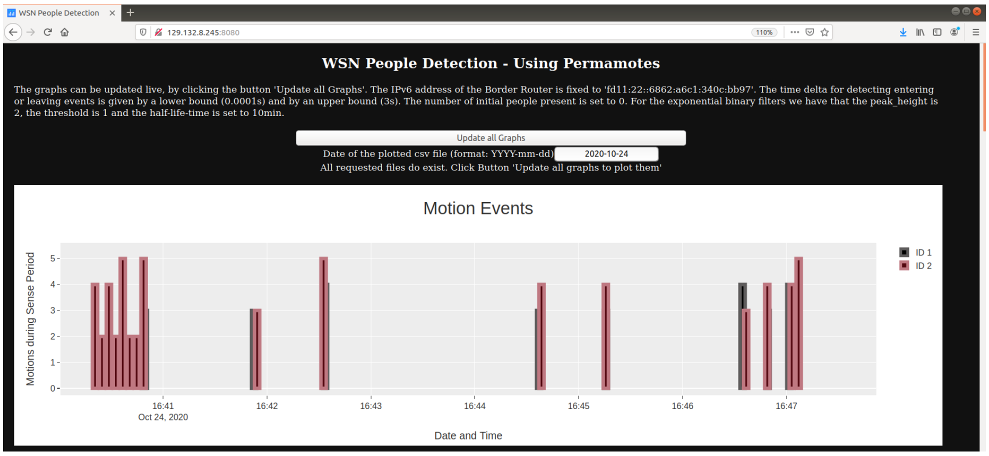


Figure 4.1: The top of the Web-GUI of the people detection system is shown, which runs on the border router at a user defined accessible IPv4 address. The part of the GUI is shown, where all motion events from both Permamotes are listed in a single plot named *Motion Events*.

4.2.1 Binary Filters

There are four different binary filters implemented, which have the goal to estimate people presence or absence with only taking the data of one PIR sensor from inside the room as input data. Three of them using an exponential approach. At that, someone is considered to be in the room if the filter value is above a certain threshold (in this work it is set to one). If the value falls below this threshold, then the room is considered to be empty. However, each of the three exponential filters handle a new motion event differently. On the contrary, all three filters have the same exponential decay rate after every peak, which in the Python program code can be set by determining the half-life (in this work it is set to ten minutes). The last filter *specific* uses another idea than exponential decays with a threshold. In the following, the four binary filters are described in more detail.

Exponential Simple

This filter, as the name suggests, is the simplest one among the three with the exponential approach. It implements the idea that lots of detected motions increases the probability that someone is staying longer in the room. Every time a new motion detection occurs, the same constant is added to the previous filter function value. That is the new peak height $f(t_k)$ after a motion detection at time t_k is calculated as

$$f(t_k) = f(t_{k-1}) \cdot exp\left(-\frac{\Delta t \cdot ln(2)}{half\_life}\right) + const \tag{4.1}$$

where const is set to two in this work. The time t_{k-1} describes the time of the last detected motion event. Moreover, $\Delta t = t_k - t_{k-1}$ describes the time difference between the last and current detected motion in minutes. Figure 4.2 visualizes this idea. The drawback of this filter is that if a lot of motion events follow each other, there is no upper bound on the function value which means it can go very long until the function value falls under the threshold after the last detected motion. This means that after lots of successive motion events the time period in which the room is assumed to be occupied can be infinitely long.

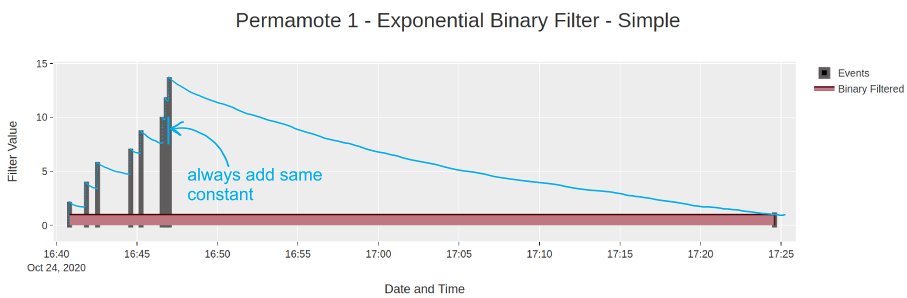


Figure 4.2: Cutout of the GUI, which shows the simple exponential binary filter with additional light blue drawing to indicate the exponential function.

Exponential Sum Last Two

This is the most complex binary filter of the four. It implements the idea that the room is at most occupied by X people, where X is set to three in this work. Hence, the exponential function values can be seen as 'probabilities' that someone is inside the room. For each possible person, we have a probability that this person is inside the room. Consequently, if $X = 3$ then, the first detected motion of the day belongs to person 1, the second to person 2, the third to person 3, the fourth detected motion again to person 1, the fifth again to person 2, etc. The 'probability' that person 1 is in the room equals its decayed exponential function value from the last motion which belonged virtually to person 1. To decide whether or not the room is occupied, all X decayed exponential function values are summed. If this value is below the threshold, the room is assumed to be empty, else the room is considered to be occupied. Figure 4.3 depicts this idea. Only the function values with the red dots are summed. Since $X = 3$, we sum the last two plus the peak of the current detected motion. In a formula, the summed exponential function value, when a new motion is detected at $t = t_k$ can be written as

$$f(t_k) = const + \sum_{n=1}^{X-1} f(t_{k-n}) \cdot exp\left(-\frac{(t_k - t_{k-n}) \cdot ln(2)}{half\_life}\right) \tag{4.2}$$

where const is set to two and X is set to three in this work. If X would be set to infinity, this filter would exactly be the same as the *exponential simple binary filter*. Therefore this approach solves the problem of the *exponential simple binary filter* that the new peak value could increase to arbitrarily large values. It gives an upper bound on the peak values, which is $const \cdot X$. This can already be seen by comparing the two small example plots in Figures 4.2 and 4.3. There, one recognizes that with the same detection data the simple filter produces already peaks of heights over 12 whereas the sum last two filter goes never above 6.

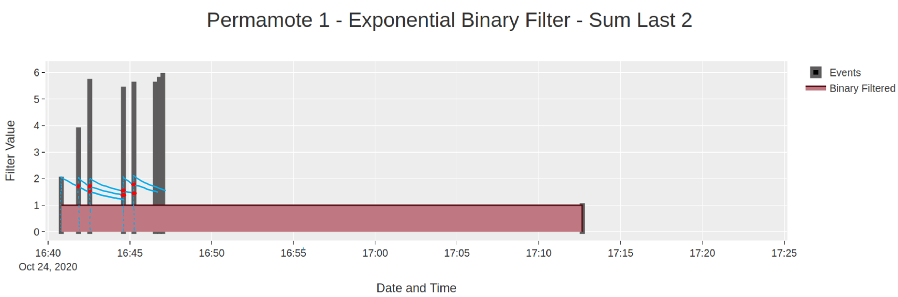


Figure 4.3: Cutout of the GUI, which shows sum last two exponential binary filter with additional drawing to indicate the working principle of the filter.

Exponential No Sum

In this approach the idea is implemented that every detected motion could be the last one. That is we have for every motion the same peak height and thus the same time until the exponential value falls under the threshold, since the decay rate stays constant. That means, no matter what remaining value the exponential function has, the new peak height at the moment of a new detected motion is always set to the same value (in this work the value is two). Figure 4.4 visualizes this idea.

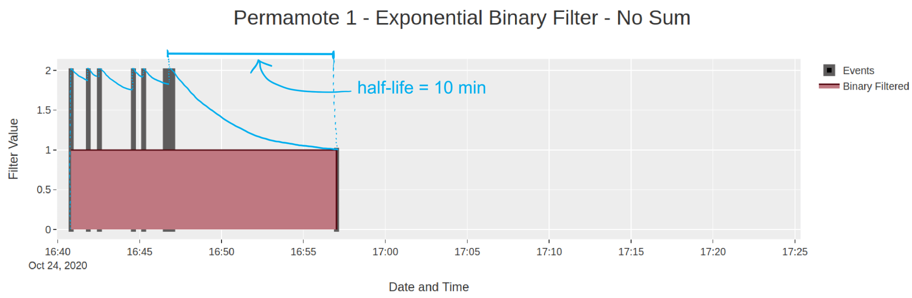


Figure 4.4: Cutout of the GUI, which shows the no sum exponential binary filter with additional light blue drawing to indicate the exponential function.

Specific

The specific binary filter does not use an exponential decaying function nor a threshold. It only looks at the time of the day when the motion is detected and starts then a timer of an user defined length. If another motion event is detected when the timer is not yet zero, then it is reset to the same initial value (see Figure 4.5 for a visualization). That means there is no summing of timer values. The initial timer values depend on the time of the day and the day itself in this work. The timer values are chosen in a heuristic manner which assumes usual working office hours and breaks. If the detected motion is on weekends, then the timer is always set to 5 minutes. If the detected motion is on a weekday and between

- 00:00 and 05:59, the timer is set to 5 minutes.
- 06:00 and 11:59, the timer is set to 30 minutes.
- 12:00 and 12:59, the timer is set to 10 minutes.
- 13:00 and 16:59, the timer is set to 30 minutes.
- 17:00 and 19:59, the timer is set to 10 minutes.
- 20:00 and 23:59, the timer is set to 5 minutes.

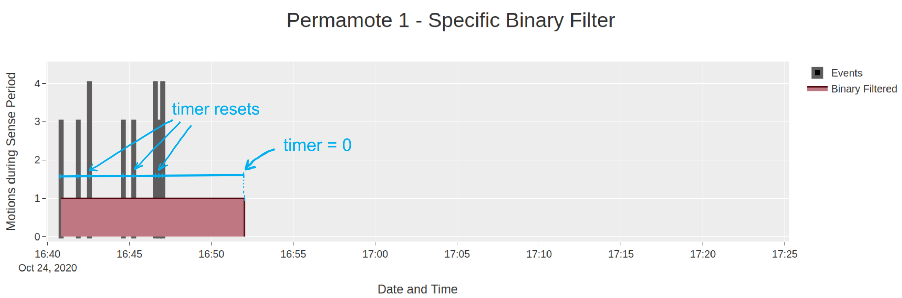


Figure 4.5: Cutout of the GUI, which shows the specific exponential binary filter with additional light blue drawing to indicate the working principle of the filter.

4.2.2 People Counting Methods

In addition to the four different binary filters (see section 4.2.1), three different ways of people counting are implemented, which are explained in this section. However, the two decision tree approaches are not implemented on the border router, but they are implemented later on while analyzing the data. Nevertheless, they could be easily transferred to the border router since they are also written in Python. Not only people counting is done with these three approaches but also binary people detection. If the counter is higher than zero, it is assumed that the room is occupied, otherwise the room is assumed to be unoccupied. The three different people counting variants are listed in the following.

List approach

This method is the most simple one and takes the detected motions from each Permamote separately as two lists as an input. Then the algorithm iterates over both lists to find all entering events and mark them with a timestamp. Having done this, it iterates again over both lists to find all leaving events. Then the two event lists were merged and sorted according to the timestamp. Finally, it iterates over the event list and adds minus (plus) one to the people counter variable for all leaving (entering) events. However, the counter can never go below zero which means if a leaving event occurs and the people counter is zero, then it stays zero and will not change to minus one. Everyday at midnight the people counter is reset to zero such that possible errors propagate at most till the end of the day. This approach can also be used for real-time people counting.

Basic decision tree approach

This method takes only a single list as input data with three different types of events. One type carries the detected motions of Permamote inside the room, another type from outside the room. The third type are reset events everyday at midnight. The input list is sorted regarding the timestamps of the entries ascending. The algorithmic idea is then to always look at two events (*last* and *next*). Then the decision tree is spanned in a sense that for each possible combination of *last* and *next* a specific action is taken. Afterwards, the *last* and *next* are updated and the new values are now the input for the decision tree. For further details see the pseudocode in the Appendix A.1. Besides this, it is important to mention that this method can be used for real-time people counting and that the counter never goes below zero.

Enhanced decision tree approach

For this approach another single list as input data is needed but this time with four different types of events. Three events are the same as in the basic decision tree approach. The fourth type is an event which indicates the last detected motion event for more than three hours from the Permamote inside the room. If such an event is seen while iterating through the list, the counter is set to zero at the time of the last detected motion (see Appendix A.2 for the pseudocode). In addition to that, it filters inside-outside-inside-outside to only inside-outside motions, if the motion events are all close to each other. The same is done for outside-inside-outside-inside motions. However, this means that it implements an improved basic decision tree approach, which only works in a retrospective manner. Thus, it is not useful for real-time people counting. Nevertheless, this approach is useful for comparing all other people counting or people detecting methods, because it should be the one with best occupancy estimation. However, since there is no data with ground truth, this statement must be seen as an educated guess for which two main explanations can be given. Firstly, it is an improvement to the basic decision tree approach, because it implements some basic filtering and has more reset events, which means that counting errors will not propagate to the end of the day but only until the last sensed motion for more than three hours of the Permamote. Secondly, this approach should also be more accurate than the binary filters because the binary filters have always some time after the detected motion event where they assume an occupied room. Consequently, during that time binary filters always have a false estimate, if the detected motion is triggered by the last person who left the room.

4.3 Limitations

It is important to mention that the deployed people detection system has its limitations due to its design. There are always errors, which can not be solved since the deployed Permamotes use PIR sensors and are deliberately designed for low power applications. Hence, it is crucial to have some weaknesses of the system in mind.

- Passive sensors as the PIR are simply triggered by changes in the environment, but they do not actively send out any kind of waves [3]. This can lead to false positive detection if e.g. a hot or cold air gust passes in front of the sensor as it could be the case if someone opens the windows. To avoid this error as often as possible the PIR sensors were placed close to the door in a way that they are only able to screen the gateway but not much more of the interior of the room respectively hallway.

- The packets, which are send by the Permamotes to the border router, use UDP as a transmission protocol. This is done, because sending packets is one of the most energy-intensive tasks a Permamote can do [3]. This means that dropped packets are lost. This error can not be avoided.

- The masking time of the used PIR sensors prevent accurate people counting. That means if more than one person walks through the observed gateway at the same time or within the masking time after each other, then this group of people would only be registered as one. This leads to an unavoidable error in the people counting. In addition to that fact another problem is that people counting errors propagate over the whole day till the next reset is made, because the current counter value is related to all previous counts since the last reset. This makes accurate people counting with this deployed low power system with PIR sensors a very hard task [3].

- Another error which can occur in the people counting application is that if the time of on one or both Permamotes are a bit out of sync (some hundredths of a second), it can happen that the timestamps of the outside and inside PIR events are messed up. Thus, in a case like that the timestamp of outside of the room would be smaller than the one from inside, even though the person actually left the room. This means that an entering event is counted instead of a leaving event. The same can happen the other way around. This limitation is taken into account by setting the synchronization interval of the Permamotes with the border router to only one minute.

- It is also very hard to have a very accurate people detection application, where it is estimated in a binary form whether or not the room is occupied. If only one sensor is used, it is hard to predict which motion is the last one and that the room is from this point in time unoccupied. Thus, it is always assumed that someone is inside the room for some time after the motion detection, which is obviously wrong if detected motion is triggered from the last person leaving the room. Hence, with the binary filters described in section 4.2.1 only an approximation of the true people presence can be made.

Since this work does not have the main goal to find the accuracy of neither a people detection nor people counting low power system with PIR sensors, it is enough to have a rough estimate, on how good the system performs for the specific environment rather than doing a deep scientific investigation about it. The same holds for all possible placements of the two Permamotes to observe the gateway. Thus, a simple test is made to get a rough estimate of accuracy of the deployed system. The set up of the deployed system is depicted in Figure 4.6. The Permamotes are placed in about 1 meter height from the floor diagonal to the entrance. The test procedure is the following. A person left the room, waited for ten seconds, entered and waited again ten seconds until the same person left again. This procedure is done 50 times. Thus, 50 leaving and 50 entering events should be detected. Actually, one leaving event was not detected, which should have been. All 99 other events were detected correctly. The reason of the missed leaving event was that a packet of the Permamote outside of the room was not received by the border router, which is also placed inside the room (see Figure 4.6). However, the waiting time of 10 seconds during this small test prevented errors due to the masking time. Consequently, it is wrong to conclude that an error of one percent can be assumed for all times based on these results. Because there might still occur some timing problems, some false positives or masking time related errors in a long run test with more than one person, which would let the error percentage increase. Additionally, one must not forget that in the people counting application an error propagates until the next reset event occurs.

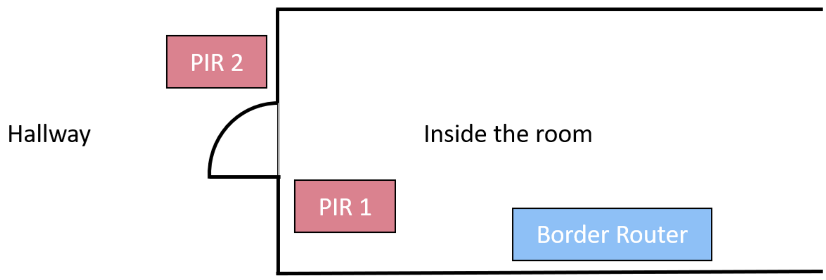


Figure 4.6: Sketch of the people counting system deployment, which is installed at the office door of the room 81 with measurement stations 16 and 17 inside (see Figure 3.1).

Chapter 5

Data Analysis

Since there is no ground truth data about room occupancy, the three chosen data analysis approaches to approximate the ground truth are described in this chapter. It is clearly specified and reasoned which time periods were chosen as inputs to the analyzes. Moreover, a brief overview of the tools used is given in this chapter in Section 5.1. All three analyzes have the goal to find an answer to the question how people occupancy and seasonal differences influence the indoor harvested PV energy. The first analysis takes whole measurement years as input (see Section 5.2). The second data examination compared the data between the corona lockdown period in Zurich in spring 2020 with the same period but of previous years (see Section 5.3). Thirdly, the time period when the people detection system was recording data in the *bright office*, was also analyzed on its own (see Section 5.4).

5.1 Tools

The data is analyzed by using different tools which are listed and shortly explained in the following.

- Principal Component Analysis (PCA) is used in order to get a qualitative statement about the influence of various factors on the amount of harvested indoor PV energy. The data is transformed in a way that the first principal component (PC) carries the biggest variance, the second PC the second greatest variance and so on [10]. 2D and 3D plots are made where the first two respectively three PCs have been plotted regarding different targets as e.g. weekdays versus weekends (see e.g. Figure 6.1b). This analysis allows only the answers 'yes, there is some influence' or 'undecided' to the question if a specific target does has an influence on the harvested energy.

- Violin Plots are another tool to describe the data. The whole distribution of the data points can be seen with violin plots [11], which makes them superior to the box plots especially if the data is multimodal (i.e. more than one peak). Additionally, they are made to compare the distributions according several targets as e.g. the harvested energy during weekdays and weekends (see Figure 6.6 as an example).

- A quantitative examination of the data is made by calculating some statistical indicators of the indoor solar harvested energy data. Moreover, a method for calculating lower and upper confidence bounds of a quantile with a user specified confidence is used [12]. This method takes no assumptions on the underlying data and its distribution.

- An augmented exponentially weighted moving average (EWMA) prediction scheme is used for the prediction of the indoor solar harvested energy with or without considering the people detection data. The standard EWMA update for the historical average [13] is given as

$$\hat{x}_t = \hat{x}_{t-1} \cdot \alpha + x_t \cdot (1 - \alpha)$$

where the hat indicates a historical average and the x without the hat indicates a measurement value. The historical average at time t is then used for the prediction for the time t+1. Additionally, the weighting factor α was set to 0.5, as it is done in [13]. The standard EWMA prediction is augmented in a way that the values $\hat{x}_t$ and $\hat{x}_{t-1}$ are not general, but case specific historical averages. If people presence data is used for the prediction, the following six different cases are considered:

  - Day time and the observed room is occupied
  - Day time and the observed room is not occupied
  - Dusk time and the observed room is occupied
  - Dusk time and the observed room is not occupied
  - Night time and the observed room is occupied
  - Night time and the observed room is not occupied.

  Only three (day, dusk and night time) cases are considered, if the people detection data is not used for the prediction.

5.2 Whole Years

For each measurement position the possible whole years are determined. Table 5.1 summarizes the analyzed years for each position. Taking only the whole years, an investigation of the measurement data regarding weekdays, weekends and seasons is possible. Since there is no ground truth during these years about the room occupancy, weekdays (people presence) and weekends (people absence) are examined to have an approximation of the ground truth. Moreover, a year is considered to be 52 weeks and not 365 days, to have only whole weeks included. The year 2020 is intentionally omitted, because of the corona lockdown and its consequences on people's presence in the rooms could have an influence on the results. Moreover, there is no whole year of data at position 16 without corona measures. Thus doing an analysis at this location in terms of the reference years is not reasonable.

The data is analyzed hourly and daily based. To get the daily based data, the harvested energy over a whole day is summed up and after that all days are compared to each other. By getting the hourly based data instead, the harvested energy is only summed over an hour.

Table 5.1: Whole years, which were considered for each measurement position.

Position	Start Date	Weekday	End Date	Weekday
06	2017-10-16	Monday	2018-10-14	Sunday
13	2017-10-16	Monday	2019-10-13	Sunday
14	2017-10-16	Monday	2019-10-13	Sunday
17	2017-10-16	Monday	2019-10-13	Sunday
18	2017-10-16	Monday	2019-10-13	Sunday

5.3 Corona Lockdown Periods

Neither for the whole years, nor for the lockdown periods there is ground truth data about the people presence. However, thanks to the corona pandemic and its measures taken by the Swiss government an interesting comparison of the data can be made, because the rooms during the lockdown period in spring 2020 were with a high certainty unoccupied. Hence, a good negative ground truth approximation is achieved during lockdown days. It is a better ground truth approximation for people presence and absence than weekdays and weekends during the whole year analysis. The lockdown in Switzerland and thus in Zurich started from Monday the 2020-03-16 and ended on Sunday the 2020-04-26 according to [14]. Due to the fact that on weekends during normal operation in previous years the rooms might also be unoccupied the whole day, only the weekdays are considered in this analysis. With this reduction on the data set, a reasonable comparison is achieved between measurement data where people are assumed to be present (pre-corona) versus absent (lockdown period in spring 2020).

For position 06 no data is available of the year 2020. Moreover, for position 16 no data of the lockdown period from previous years was recorded. In addition to that the data for position 18 can not be considered for the lockdown analysis, because some change in lighting conditions happened in fall 2019. This makes data comparison between 2020 and previous years unnecessary. Hence, Table 5.2 shows the data traces, which are taken as input for the corona lockdown periods analysis for the remaining positions 13, 14 and 17. These data traces are examined on a daily and hourly basis.

Table 5.2: Corona lockdown periods, which are considered for each location without taking the weekends into account. For each year we have exactly six weeks which means 30 days per year are taken. That is 60 days of normal operation versus 30 days of lockdown are analyzed.

Start Date	End Date	Presence of People
2018-03-19	2018-04-29	mostly present
2019-03-18	2019-04-28	mostly present
2020-03-16	2020-04-26	mostly absent

5.4 People Detection Period

For the people detection analysis the time between 2020-10-23 and 2020-11-28 is analyzed. However, there is no ground truth data about the people presence during this time. Thus, the data from the low power system needs to be seen as an approximation of the ground truth. On one hand there is the data of the measurement stations as it is for the lockdown periods as well as for the whole years analysis. On the other hand people counting as well as people detection data from the deployed low power system with PIR sensors is available (see Section 4). These two data traces are mixed together and the following points are examined.

- The different people detection methods are compared between each other. Since there is no ground truth about the people presence in the natural light driven room 81 (see 3.1) during this period, their room occupancy estimates are compared during night times only, i.e. times between sunset and sunrise. Based on the assumption that the people present always turn on the lights, the comparison between them is done. The idea is that, if people are present during night times, more energy is harvested during this time compared to unoccupied times because of the switched-on lights. However, if people are absent, there is no or very little light and thus no or very little energy harvested. Consequently, the following scale is introduced. The higher the mean of the harvested energy during the occupied room estimations of a specific detection method, the better is the method.

- The absolute number of people inside the room is examined regarding the harvested energy. Specifically, it is investigated, if it is true that more energy is harvested when two (or more) persons are inside a room compared to only one person being present.

- The indoor solar harvested energy regarding room occupancy is analyzed on a hourly basis.

- Two different prediction schemes are evaluated. One scheme uses the people detection data, the other does not. The harvested energy is predicted for the next 5 minutes, 10 minutes and 1 hour.

Chapter 6

Results

Since there is no ground truth data about the room occupancy, three approximation of it were examined for each position as it is described in Section 3. Even though one approximation of the ground truth with respect to the room occupancy might be better than another, the results of each approximation are shown for each position. Nevertheless, the supposedly best available approximations with respect to room occupancy is for the position 06 and 18 the comparison between weekdays (people present) and weekends (people absend). These results can be found in Section 6.2. For positions 13 and 14 it is the comparison between pre-corona (people present) and corona (people absent) and its results can be found in Section 6.3. For position 16 and 17 the supposedly best available approximation is achieved by the low power detection system of which the results can be found in Section 6.4.

At the beginning of this chapter, the results of the principal component analyzes of each measurement station are shown (see Section 6.1). In Sections 6.2 and 6.3 the results of the quantitative statistical data analysis of the ETH indoor solar energy harvested data set are shown for each measurement station separately. In addition to that, the results of the indoor harvested PV energy measurement analysis with taking the data of the deployed people detection system into account are shown in Section 6.4.

Only the most interesting results of each position is shown. All the statistical indicators and all kind of different plots can be found in the project folder *sa_kiserm/workspace/post_process_eth_data_set*.

6.1 Qualitative Investigation with PCA

The different rooms are analyzed qualitatively by conducting a principal component analysis (PCA) with respect to the influence of seasons and people presence. The detailed specifications of the measurement stations and their rooms is listed in Table 3.1. For each measurement station, two PCA plots are shown, such that the two following questions can be answered qualitatively.

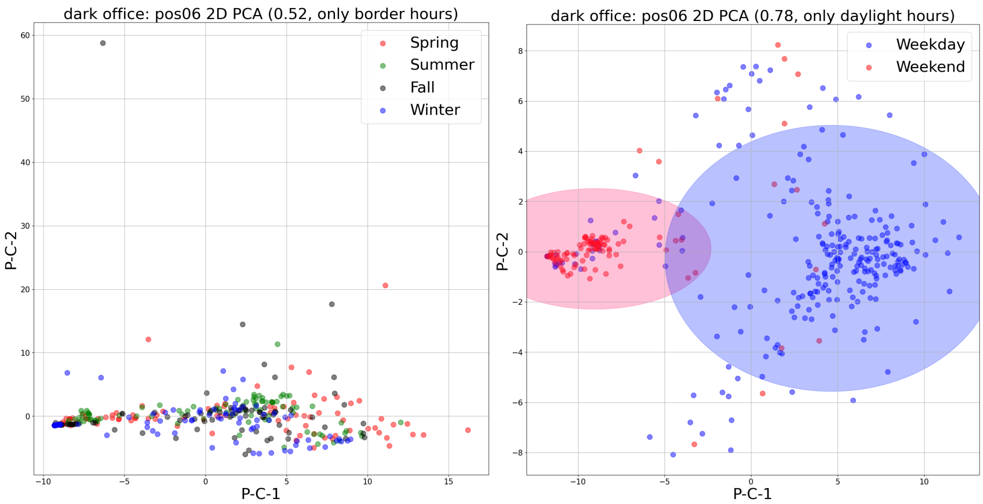
1. Do the four seasons have an influence on the amount of harvested energy?
2. Does people presence have an influence on the amount of harvested energy?

For both of this questions, the answer is either yes or undecided. The answers are found by looking at the clustering of the data. Every data point in the PCA plots corresponds to all considered energy measurements of one day. At each ten minutes the energy is measured. The numbers of features for one day is the number of the considered ten minutes energy measurements as well as one additional feature. This additional feature is the mean of all considered features of that day. If for example the measurements between 10am and 3pm are considered in the PCA, then 37 features describe one day. Moreover, the shown red and blue ellipses are drawn to indicate how much the blue and red data points differ. They are centered at the medians of the first and second PC. Moreover, the width of an ellipse is the distance between the 95% and 5% quantile of the first PC whereas the height is the the distance between the 95% and 5% quantile of the second PC. For the seasonal PCA plot only border hours are considered in the PCA, which describes hours from 6am until 9am and from 4pm until 8pm. Because it is assumed that during these hours of the day, the effect of the seasons is best observable. To answer the people presence question, only the hours 10am until 3pm are considered for the PCA. The assumption here is that most of the people show up during these hours. Thus, late and early office hours are not considered.

In the following the most interesting PCA plots with respect to room occupancy and seasonal differences are shown for each measurement position separately on a single page. This way, it is easier for the reader to link the text with the corresponding figures.

Position 06 - Dark office

Looking at the Figure 6.1a no clustering regarding the different season can be observed. Moreover, the first two principal components only include 52% of the overall variance. Hence, the answer to the first question is undecided. Looking at the figure 6.1b there is a clear separation of the weekend (red) and weekdays (blue) data points. Moreover, this 2D PCA plot shows 78% of the underlying overall variance. Assuming that for most of the daylight hours the office is occupied during weekdays and not during weekends, then it can be concluded based on this PCA plot, that people presence has an influence on the amount of harvested energy for this measurement station.

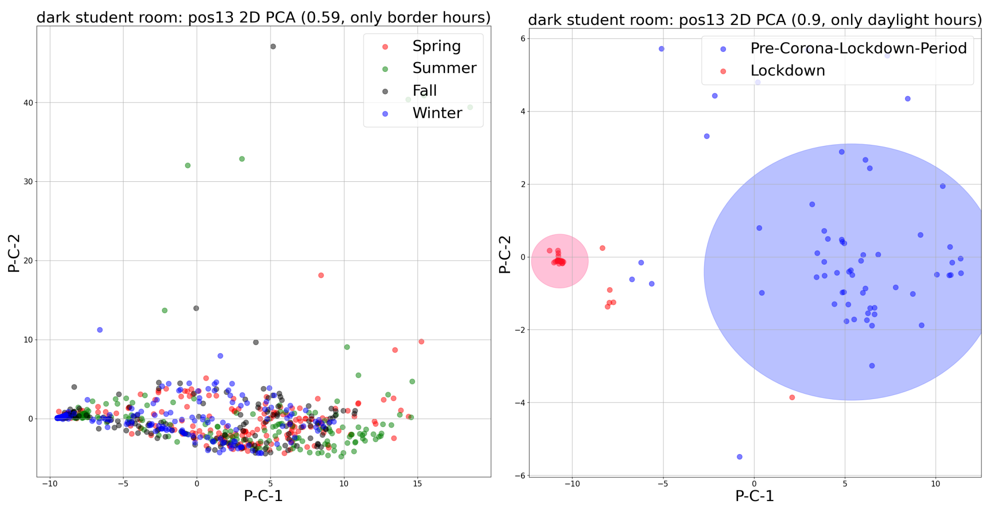


(a) PCA plot season (b) PCA plot weekend

Figure 6.1: PCA plots of measurement station 06.

Position 13 - Dark Student Room

For measurement station 13 the PCA plot to visualize, if there is clustering observable or not regarding the different season, is shown in Figure 6.2a. The first two principal components include 59% of the overall variance. However, no clustering or separation between the seasonal data points can be seen. Thus, the answer to the first question is undecided. On the other hand, the answer to the second question is yes. A clear separation and clustering of pre-corona (blue) and corona (red) lockdown periods can be observed in Figure 6.2. The first two principal components already include 90% of the underlying overall variance. Since lockdown days (red) represent days of unoccupied rooms and pre-corona days (blue) stand for occupied rooms, it can be concluded that people presence has an effect on the amount of harvested energy.



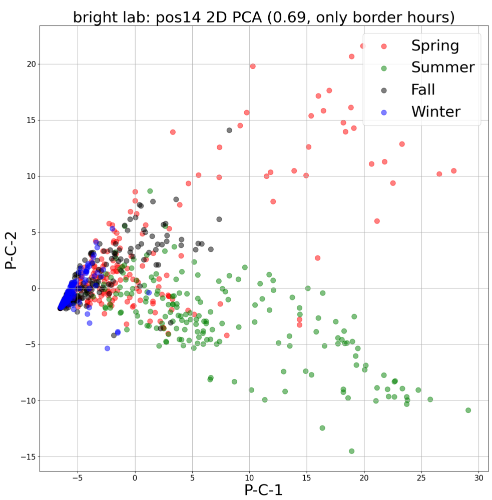
(a) PCA plot season

(b) PCA plot lockdown

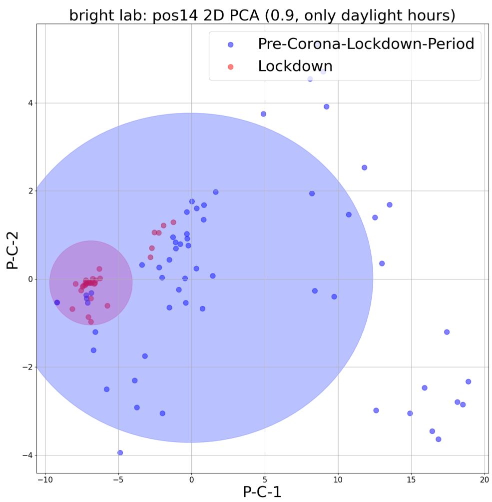
Figure 6.2: PCA plots of measurement station 13.

Position 14 - Bright Lab

The winter data points (blue) are very clustered in the Figure 6.3a. Moreover, the summer data points (green) are quite separated from the winter (blue) data points. However, the spring (red) and fall (black) data points do not lie in different clusters. However, the first two principal components include only 69% of the overall variance. The seasonal difference between winter and summer or winter and spring can affect the amount of harvested energy, but it is not crystal clear from this PCA plot. Especially only 69% of the variance is shown in this plot. Hence, it can be concluded that there is a tendency that seasons can have an effect on the amount of indoor harvested PV energy. On the other hand, looking at Figure 6.3b a red cluster of the lockdown data points can be seen. However, there is no clear separation of the data points regarding corona (red) and pre-corona (blue) days. These first two principal components already include 90% of the overall underlying variance. Thus, no statement can be made about the effect of people presence on the amount of harvested energy based on this plot.



(a) PCA plot season

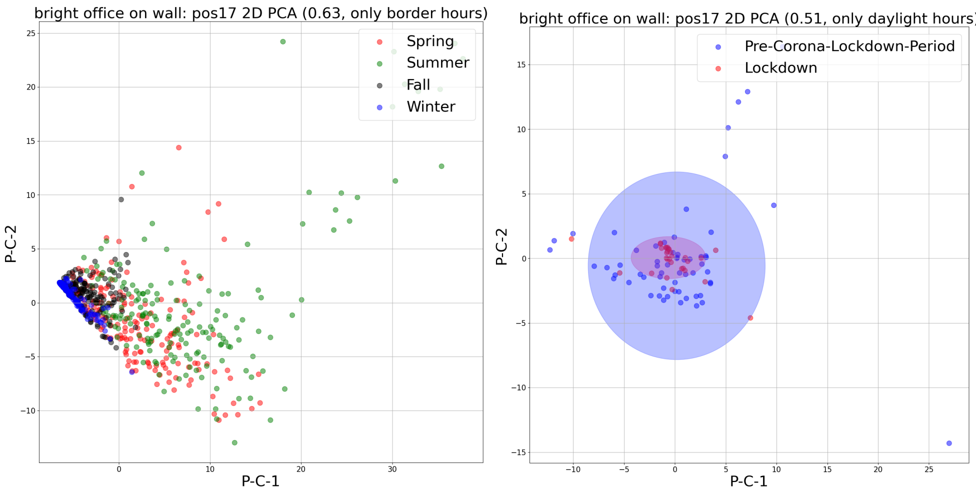


(b) PCA plot lockdown

Figure 6.3: PCA plots of measurement station 14.

Position 17 - Bright office on wall

Figure 6.4a shows a clustering of the different seasons. Even tough the clustering is observable, the included variance in the first two principal components is only 63%. Consequently, there might be an seasonal influence on the amount of harvested energy, but it is not very clear, because lots of the underlying variance is missing in this plot. The same statement regarding the depicted variance holds also for Figure 6.4b, where only 51% of the overall variance is depicted. However, there are no two separated clusters of pre-corona and corona data points. Thus, there is no answer to the people presence influence on the harvested energy based on this PCA plot.

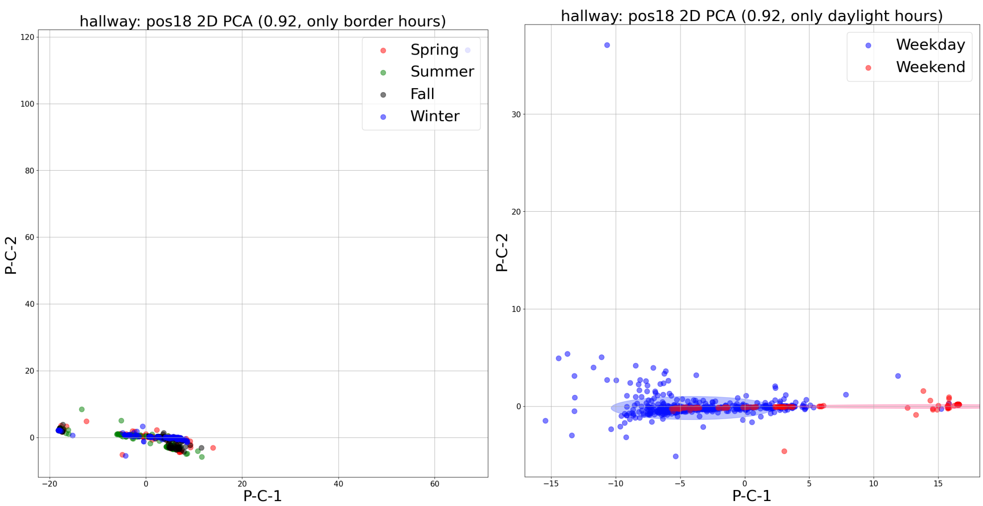


(a) PCA plot season (b) PCA plot lockdown

Figure 6.4: PCA plots for measurement station 17.

Position 18 - Hallway

There is no answer if the seasons affect the amount of harvested energy at the measurement station 18. The data points in the Figure 6.5a are neither clearly clustered, nor separated according to the different seasons. On the other side, the most data points of the weekdays (blue) are on the left hand side where the data points of the weekends (red) are mostly on the right hand side. Thus a slight clustering can be seen. Moreover, the first two principal components include already 92% of the overall underlying variance. Nevertheless, a clear separation between weekday and weekend data points is not depicted. Hence, it is concluded based on this PCA plot that it is undecided if people presence affects the amount of harvested energy at measurement station 18 or not.



(a) PCA plot season (b) PCA plot weekend

Figure 6.5: PCA plots for measurement station 18.

6.2 Whole Years

The underlying data traces for each position is specified in section 5.2. Each position is examined on a daily and hourly basis regarding the targets

- Weekend
  - weekday $\Longleftrightarrow$ Monday, Tuesday, ..., Friday
  - weekend $\Longleftrightarrow$ Saturday, Sunday
- Season
  - spring $\Longleftrightarrow$ March, April, Mai
  - summer $\Longleftrightarrow$ June, July, August
  - fall $\Longleftrightarrow$ September, October, November
  - winter $\Longleftrightarrow$ December, January, February

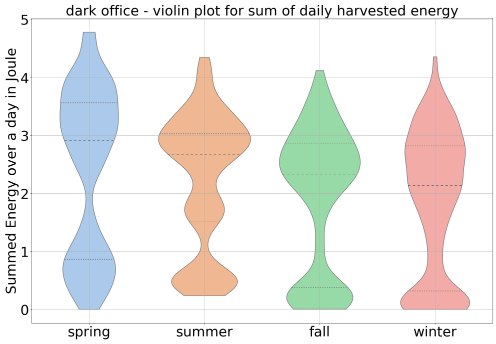
In the following the most interesting plots and statistical indicators are shown for each measurement position separately. The results of the whole years analysis of position 06, 13, 14, 17 and 18 are shown in the following. At each position the results of the daily based analysis are shown first, then the hourly based results are described. At the end of this section a brief summary of the most important quantitative daily based results is given (see Section 6.2.6).

6.2.1 Position 06 - Dark Office

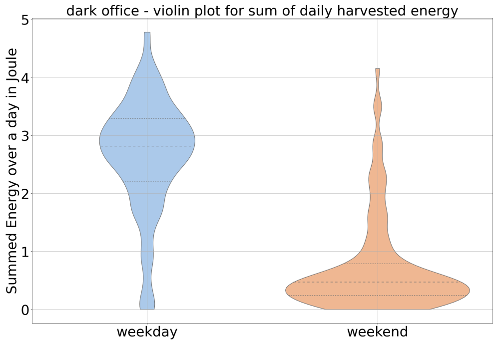
Daily Based Analysis

Figure 6.6b shows the respective distributions of weekdays (blue) and weekends (orange) of the indoor solar harvested energy summed over a day as a violin plot where the three lines show the quartiles. The statistical analysis shows that the median of the daily harvested energy during weekdays is significantly ($\alpha = 0.05$) higher compared to the median of weekends (see Table 6.1). The median of weekends is 81.4% smaller than the median of the weekdays with a confidence of 95%. Moreover, the interquartile ranges of weekdays and weekends do not overlap with a confidence of 95%.

On the contrary, the violin plot of the four seasons (see Figure 6.6a) show different shapes of the violins as well as different medians. Specifically, the median of the daily harvested energy in spring is significantly ($\alpha = 0.05$) higher than in winter. However, the median in winter is only 2.87% smaller than in spring with a confidence of 95%. All other seasonal combinations do not have disjoint 95% confidence intervals of the median values.



(a) Target season



(b) Target weekend

Figure 6.6: Daily based violin plots of the measurement station 06.

Table 6.1: Quantile values of the daily harvested energy in Joule of position 06 for weekdays and weekends of one whole year of measurements between 2017-10-16 and 2018-10-14. The lower and upper bounds show an interval, which includes the true quantile value with a confidence of 95%.

Property	**Weekday**	**Weekend**
25% quantile lower bound	2.07 J	0.13 J
25% quantile	**2.2 J**	**0.24 J**
25% quantile upper bound	2.37 J	0.3 J
Median lower bound	2.72 J	0.33 J
Median	**2.81 J**	**0.47 J**
Median upper bound	2.90 J	0.51 J
75% quantile lower bound	3.2 J	0.65 J
75% quantile	**3.29 J**	**0.79 J**
75% quantile upper bound	3.39 J	1.2 J
Number of days	260	104

Hourly Based Analysis

In Figure 6.7 the influence of the people presence is depicted for each hour. From 8am until 8pm the median of the harvested energy during weekdays is significantly higher than during weekends. The seasonal influence at this measurement station is only at 6am, 7am, 7pm and 8pm observable regarding a significance difference in the median of the hourly harvested energy (see Figure 6.8).

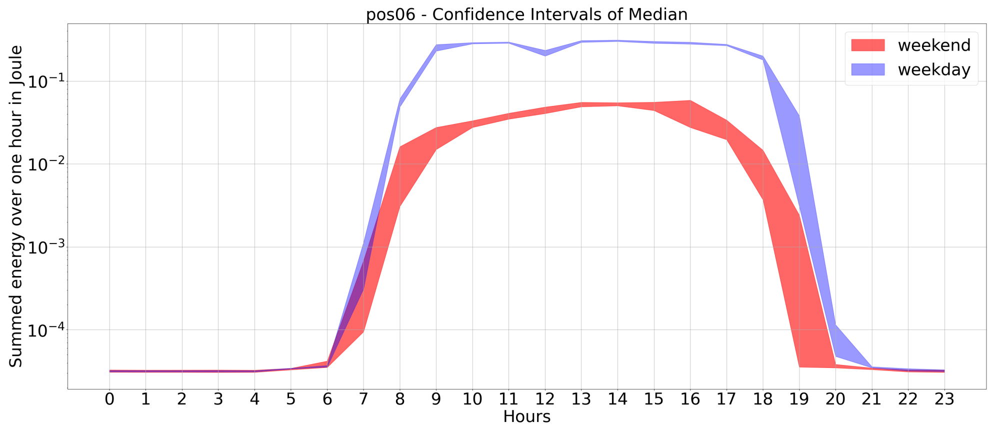


Figure 6.7: Hourly based plot of the measurement station 06 regarding weekdays (blue) and weekends (red). The colorized bands are the 95% confidence intervals of the hourly harvested indoor PV energy median.

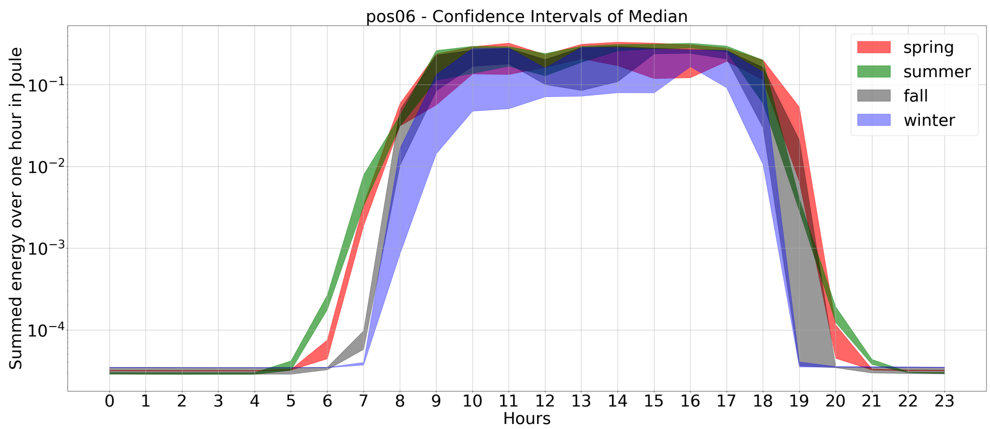
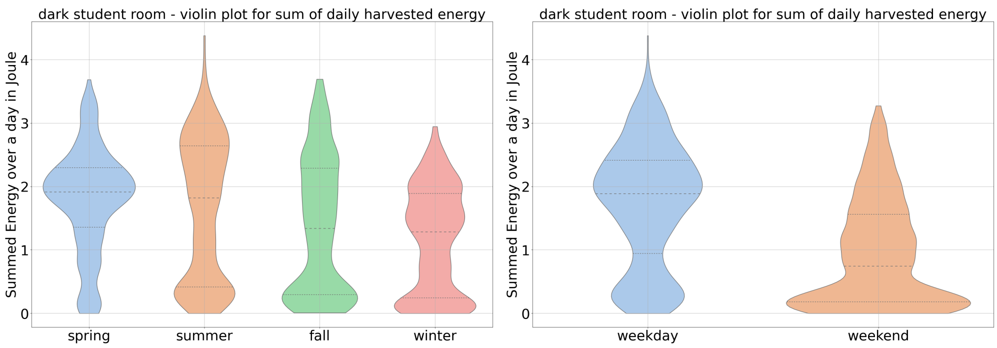


Figure 6.8: Hourly based plot of the measurement station 06 regarding the four seasons (spring [red], summer [green], fall [black] and winter [blue]). The colorized bands are the 95% confidence intervals of the hourly harvested indoor PV energy median.

6.2.2 Position 13 - Dark Student Room

Daily Based Analysis

By analyzing the targets weekend and season as well as by taking the summed harvested energy over a day as input, the medians of weekdays and weekends, winter and summer as well as winter and spring are significantly different ($\alpha = 0.05$). The median of the harvested energy during weekends is 48.9% smaller than during weekdays with a confidence of 95%. Besides this, the median of the harvested energy in winter is 22.6% smaller than in spring and 7.1% smaller than in summer with a confidence of 95%. The numbers are listed in Table 6.2. The interquartile ranges of weekdays and weekends and all the different seasons are overlapping with a confidence of 95%, as it is visualized in Figure 6.9.



(a) Target season (b) Target weekend

Figure 6.9: Daily based violin plots of the measurement station 13.

Table 6.2: Median values of the daily harvested energy in Joule of position 13 for weekdays and weekends of two whole year of measurements between 2017-10-16 and 2019-10-13. The lower and upper bounds show an interval, which includes the true median value with a confidence of 95%.

Property	**Weekday**	**Weekend**	**Winter**	**Summer**	**Spring**
Median LB	1.82 J	0.44 J	1.03 J	1.55 J	1.86 J
Median	**1.89 J**	**0.74 J**	**1.28 J**	**1.82 J**	**1.91 J**
Median UB	1.96 J	0.93 J	1.44 J	2.09 J	2.00 J
Number of days	520	208	180	184	184

Hourly Based Analysis

In Figure 6.10 the influence of the people presence is depicted for each hour. From 7am until 7pm the median of the harvested energy during weekdays is significantly higher than during weekends. The seasonal influence at this measurement station is from 5am until 8am observable regarding the significance difference in the median of the hourly harvested energy (see Figure 6.11).

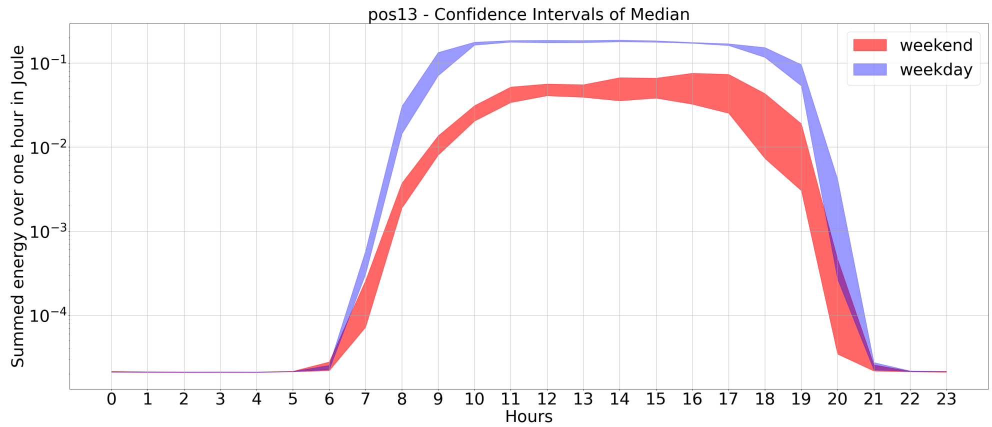


Figure 6.10: Hourly based plot of the measurement station 13 regarding weekdays (blue) and weekends (red). The colorized bands are the 95% confidence intervals of the hourly harvested indoor PV energy median.

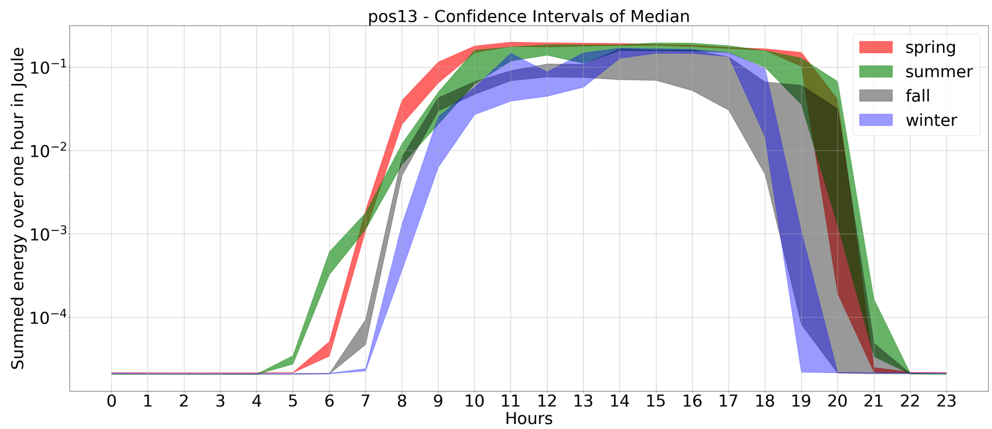


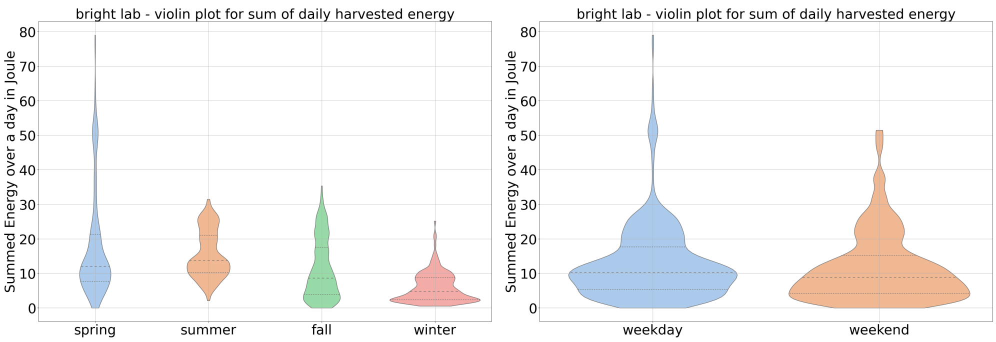
Figure 6.11: Hourly based plot of the measurement station 13 regarding the four seasons (spring [red], summer [green], fall [black] and winter [blue]). The colorized bands are the 95% confidence intervals of the hourly harvested indoor PV energy median.

6.2.3 Position 14 - Bright Lab

Daily Based Analysis

Analyzing whole years taking the summed harvested energy over a day as input yields to the fact that the median values of weekdays and weekends differ statistically significantly ($\alpha = 0.05$). However, the median of the daily harvested energy during weekends is only 0.3% smaller than during weekdays with a confidence of 95%. The median values are 10.29 J for weekdays (lower bound 9.86 J, upper bound 11.06 J) and 8.85 J for weekends (lower bound 7.5 J, upper bound 9.83 J).

On the contrary, the differences regarding the four seasons are bigger (see Table 6.3). The median values of the daily harvested energy in spring and fall, spring and winter, summer and fall, summer and winter as well as fall and winter differ statistically significantly. The median in fall is 11.6% smaller than in spring and 22% smaller than in summer with a confidence of 95%. The median in winter is is 26% smaller than in fall, 52.1% smaller than in spring and 57.8% smaller than in summer with a confidence of 95%. Besides this, the interquartile ranges of weekdays and weekends overlap with a confidence of 95% as it is visualized in the Figure 6.12b. The seasonal difference in the amount of the daily harvested energy between summer and winter is so big that the intersection of the interquartile ranges is empty (see Figure 6.12a) with a confidence of 95%. The corresponding numbers are listed in Table 6.3.



(a) Target season (b) Target weekend

Figure 6.12: Daily based violin plots of the measurement station 14.

Table 6.3: Quantile values of the daily harvested energy in Joule of position 14 for all seasons based on the data of two whole year of measurements between 2017-10-16 and 2019-10-13. The lower and upper bounds show an interval, which includes the true quantile value with a confidence of 95%.

Property	**Spring**	**Summer**	**Fall**	**Winter**
25% quantile lower bound	6.57 J	9.39 J	2.86 J	2.13 J
25% quantile	**7.69 J**	**10.23 J**	**3.87 J**	**2.36 J**
25% quantile upper bound	9.09 J	11.0 J	4.61 J	2.75 J
Median lower bound	11.32 J	12.83 J	7.32 J	4.07 J
Median	**12.03 J**	**13.73 J**	**8.62 J**	**4.75 J**
Median upper bound	12.93 J	14.46 J	10.01 J	5.42 J
75% quantile lower bound	18.37 J	19.23 J	15.67 J	7.76 J
75% quantile	**21.3 J**	**21.05 J**	**17.49 J**	**8.73 J**
75% quantile upper bound	25.62 J	23.57 J	20.82 J	9.17 J
Number of days	184	184	180	180

Hourly Based Analysis

In Figure 6.13 it is depicted that the there is no influence of the people presence for each hour with the ground truth approximation of weekend versus weekdays. However, the seasonal influence at this measurement station is the whole day observable regarding the significance difference in the median of the hourly harvested energy (see Figure 6.14).

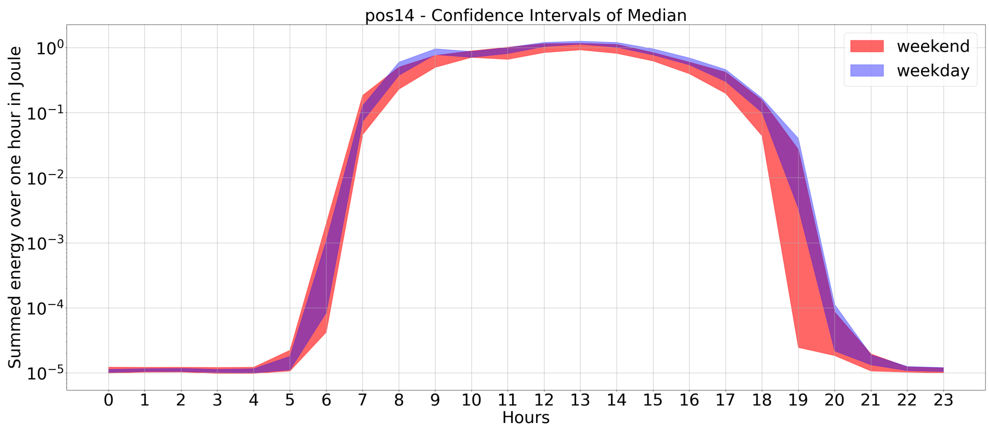


Figure 6.13: Hourly based plot of the measurement station 14 regarding weekdays (blue) and weekends (red). The colorized bands are the 95% confidence intervals of the hourly harvested indoor PV energy median.

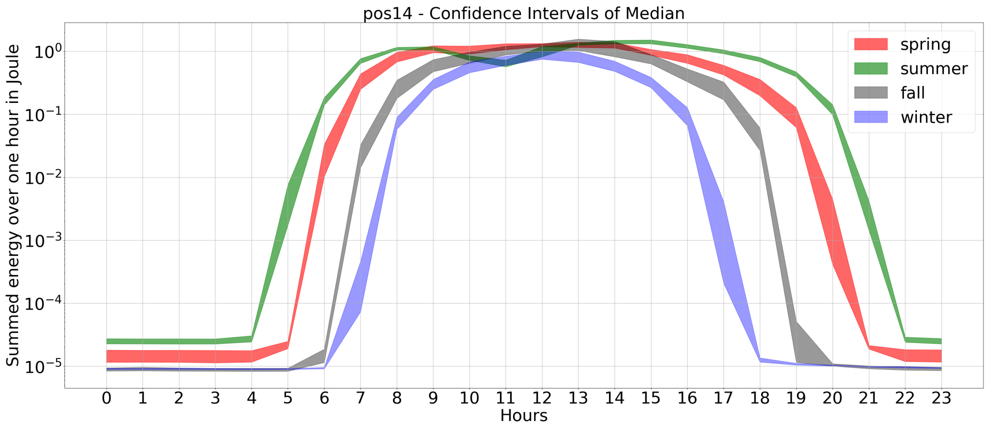


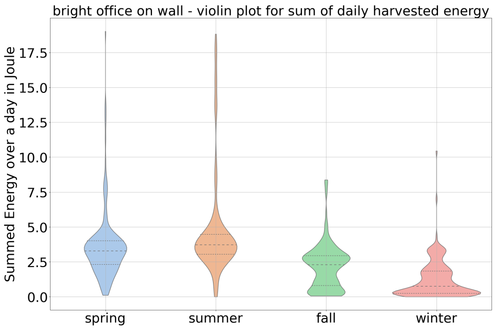
Figure 6.14: Hourly based plot of the measurement station 14 regarding the four seasons (spring [red], summer [green], fall [black] and winter [blue]). The colorized bands are the 95% confidence intervals of the hourly harvested indoor PV energy median.

6.2.4 Position 17 - Bright Office on Wall

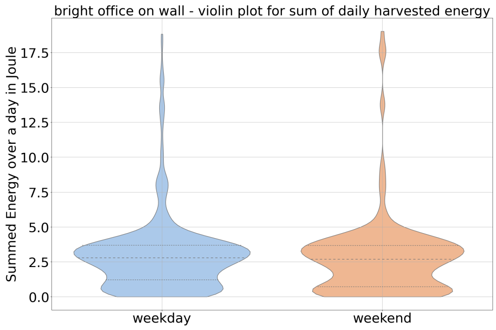
Daily Based Analysis

Analyzing the reference years taking the summed harvested energy over a day as input yields to the fact that the median values of weekdays and weekends have no disjoint 95% confidence intervals. The median values are 2.79 J for weekdays (lower bound 2.6 J, upper bound 2.89 J) and 2.69 J for weekends (lower bound 2.32 J, upper bound 2.98 J). Moreover, the interquartile ranges of weekdays and weekends overlap with a confidence of 95% as it is visualized in the Figure 6.15b.

The 95% confidence intervals of the median values of the daily harvested energy in spring, summer, fall and winter are disjoint to each other. The median in spring is 2% smaller than in summer with a confidence of 95%. The median in fall is 16.4% smaller than in spring and the median in winter is 53.4% smaller than in fall with a confidence of 95%. The seasonal differences between summer and winter are so big that the interquartile ranges do not overlap (see Figure 6.12a) with a confidence of 95%. In Table 6.4 the quantiles of the daily harvested energy of each season are listed.



(a) Target season



(b) Target weekend

Figure 6.15: Daily based violin plots of the measurement station 17.

Table 6.4: Quantile values of the daily harvested energy in Joule of position 17 for all seasons of two whole year of measurements between 2017-10-16 and 2019-10-13. The lower and upper bounds show an interval, which includes the true quantile value with a confidence of 95%.

Property	Spring	Summer	Fall	Winter
25% quantile lower bound	1.89 J	2.9 J	0.45 J	0.21 J
25% quantile	**2.31 J**	**3.04 J**	**0.8 J**	**0.25 J**
25% quantile upper bound	2.49 J	3.21 J	1.13 J	0.28 J
Median lower bound	3.11 J	3.54 J	2.06 J	0.55 J
Median Quantile	**3.28 J**	**3.72 J**	**2.29 J**	**0.75 J**
Median upper bound	3.47 J	3.84 J	2.60 J	0.96 J
75% quantile lower bound	3.8 J	4.17 J	2.85 J	1.65 J
75% quantile	**4.02 J**	**4.46 J**	**2.94 J**	**1.84 J**
75% quantile upper bound	4.29 J	4.94 J	3.26 J	2.1 J
Number of days	184	184	180	180

Hourly Based Analysis

In Figure 6.16 it is depicted that the there is no influence of the people presence for each hour with the ground truth approximation of weekend versus weekdays. However, the seasonal influence at this measurement station is from 5am until 9pm observable regarding the significance difference in the median of the hourly harvested energy (see Figure 6.17).

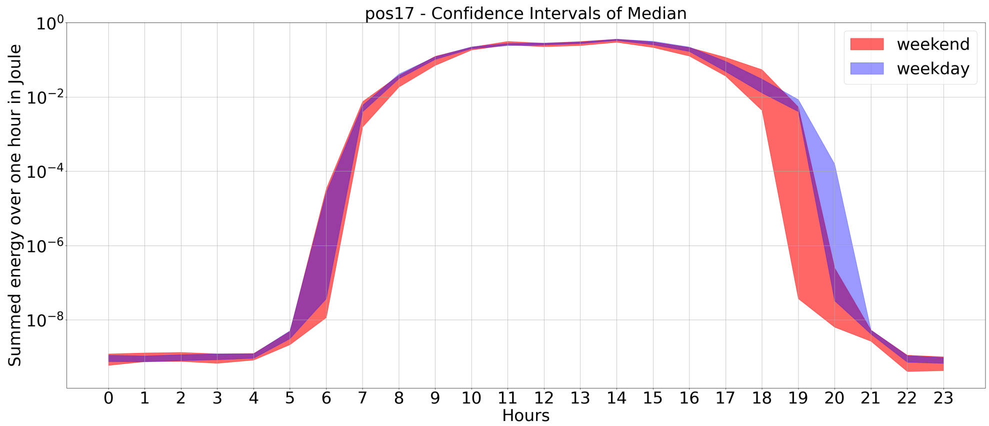


Figure 6.16: Hourly based plot of the measurement station 17 regarding weekdays (blue) and weekends (red). The colorized bands are the 95% confidence intervals of the hourly harvested indoor PV energy median.

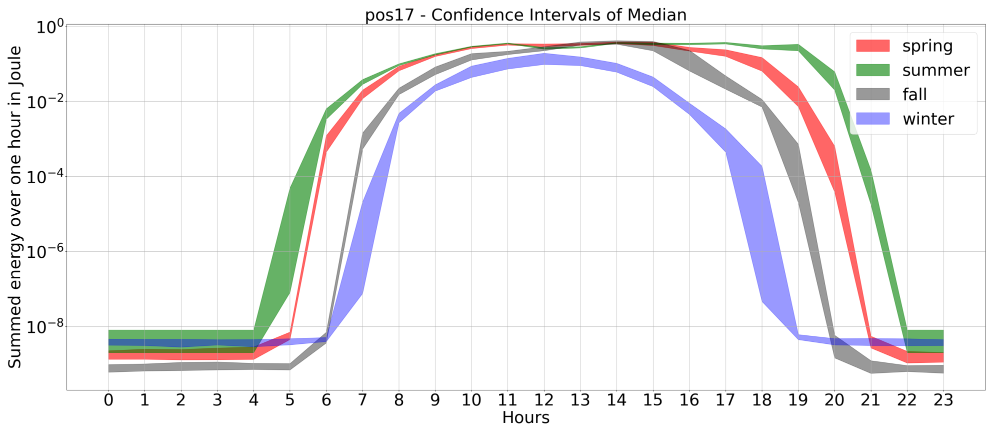
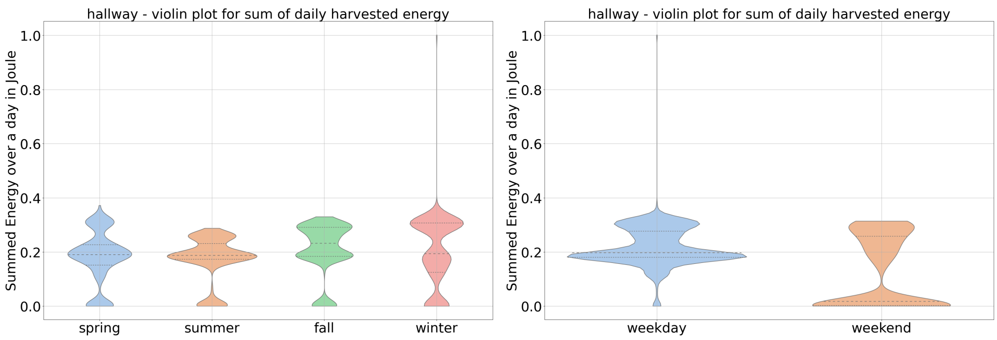


Figure 6.17: Hourly based plot of the measurement station 17 regarding the four seasons (spring [red], summer [green], fall [black] and winter [blue]). The colorized bands are the 95% confidence intervals of the hourly harvested indoor PV energy median.

6.2.5 Position 18 - Hallway

Daily Based Analysis

Examining the whole years data of measurement station 18 regarding the summed harvested energy over a day yields to the understanding that the seasonal differences regarding the median values and interquartile ranges are small. Hence, as it is visualized in the Figure 6.18a, the intersection of the interquartile ranges of any season pair is not empty with a confidence of 95%. Moreover, only the 95% confidence interval of the median values in summer and fall are disjoint. However, the median in summer is only 0.003% smaller than in fall with a confidence of 95%. Similarly, weekdays and weekends show a significance difference regarding their median values, but the median of the daily harvested energy on weekends is only 0.005% smaller than on weekdays with a confidence of 95%. All corresponding values are listed in Table 6.5.



(a) Target season (b) Target weekend

Figure 6.18: Daily based violin plots of the measurement station 18.

Table 6.5: Median values of the daily harvested energy in Joule of position 18 for weekdays and weekends of two whole year of measurements between 2017-10-16 and 2019-10-13. The lower and upper bounds show an interval, which includes the true quantile value with a confidence of 95%.

Property	**Weekday**	**Weekend**	**Summer**	**Fall**
Median LB	0.192 J	0.006 J	0.183 J	0.197 J
Median	**0.197 J**	**0.018 J**	**0.187 J**	**0.232 J**
Median UB	0.201 J	0.182 J	0.191 J	0.238 J
Number of days	520	208	184	180

Hourly Based Analysis

In Figure 6.19 it is depicted that the there is some influence on a hourly basis of the people presence on the harvested energy with the ground truth approximation of weekend versus weekdays. During the hours 6am until 8pm significantly more energy is harvested during weekdays than during weekends. However, the seasonal influence at this measurement station is never really observable in the Figure 6.20 regarding the significance difference in the median of the hourly harvested energy.

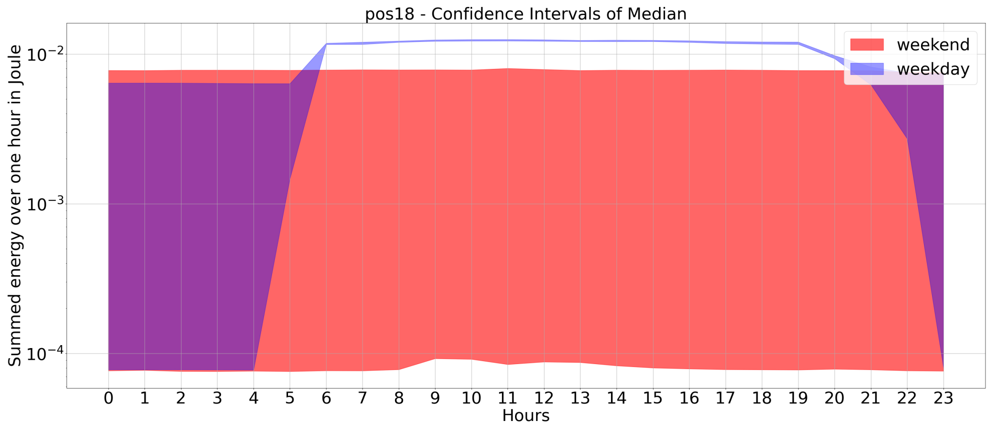


Figure 6.19: Hourly based plot of the measurement station 18 regarding weekdays (blue) and weekends (red). The colorized bands are the 95% confidence intervals of the hourly harvested indoor PV energy median.

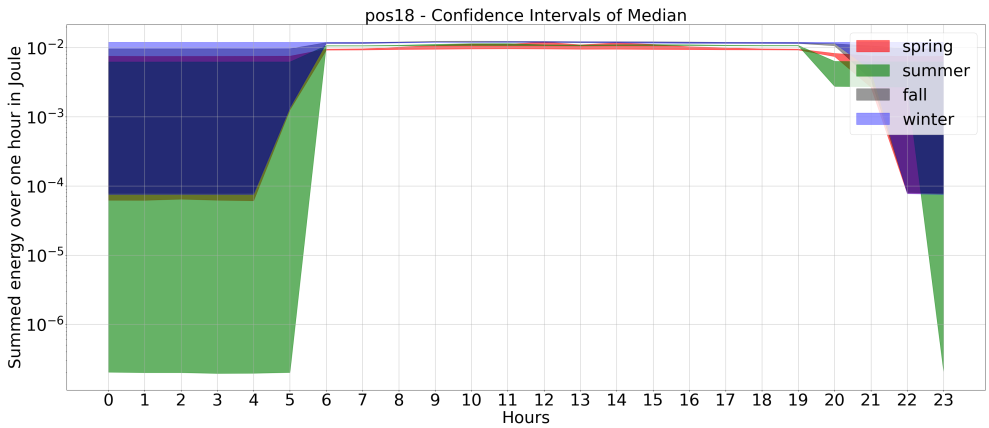


Figure 6.20: Hourly based plot of the measurement station 18 regarding the four seasons (spring [red], summer [green], fall [black] and winter [blue]). The colorized bands are the 95% confidence intervals of the hourly harvested indoor PV energy median.

6.2.6 Summary of the daily based results

Table 6.6 summarizes all the answers based on the daily based analysis of the whole years data for each position to the questions:

1. Are the 95% confidence intervals of the different pairwise median values disjoint? If yes, the relative difference is given. If no, the combination of the corresponding targets are not listed. → column *rel. dif. in median* in table 6.6

2. Is the intersection of the interquartile ranges of the pairwise different target value distributions an empty set with a confidence of 95%? → column *Middle 50%* in table 6.6

Table 6.6: Summary of the most important statistical findings of the whole years daily based analysis for each measurement position. The target 1 has always significantly higher median of the daily harvested energy compared to the median of target 2.

Position	**Target 1**	**Target 2**	**rel. dif. in median**	**Middle 50%**
06	weekday	weekend	-81.4%	✓
06	spring	winter	-2.87%	✗
13	weekday	weekend	-48.9%	✗
13	spring	winter	-22.6%	✗
13	summer	winter	-7.1%	✗
14	weekday	weekend	-0.3%	✗
14	spring	fall	-11.6%	✗
14	spring	winter	-52.1%	✗
14	summer	fall	-22%	✗
14	summer	winter	-57.8%	✓
14	fall	winter	-26%	✗
17	summer	spring	-2%	✗
17	summer	fall	-26.55%	✗
17	summer	winter	-72.88%	✓
17	spring	fall	-16.4%	✗
17	spring	winter	-69.13%	✗
17	fall	winter	-53.4%	✗
18	weekday	weekend	-0.005%	✗
18	summer	fall	-0.003%	✗

6.3 Lockdown Period

For each measurement station a daily and hourly based analysis of the underlying lockdown periods data traces (see Section 5.3) is conducted. For each station 13, 14 and 17 the results are shown regarding the only target lockdown, which differentiates between

- corona $\Longleftrightarrow$ data from 2020-03-16 till 2020-04-26 without weekends
- pre-corona $\Longleftrightarrow$ data from 2018-03-19 till 2018-04-29 and from 2019-03-18 till 2019-04-28 without weekends.

For the rest of this work it holds that the term *pre-corona* refers to the before mentioned 60 days of data, whereas *corona* corresponds to the before specified 30 days of data. Moreover, the term lockdown periods refers to both *pre-corona* and *corona*.

In the following the most interesting plots and statistical indicators are shown for each measurement position separately. The results of the lockdown period analysis of position 13, 14 and 17 are shown in the following. At each position the results of the daily based analysis are shown first, then the hourly based results are described. At the end of this section a brief summary of the most important quantitative daily based results is given (see Section 6.3.4).

6.3.1 Position 13 - Dark Student Room

Daily Based Analysis

By analyzing the target *lockdown* more deeply and taking the summed harvested energy over a day as input, it can be concluded that the medians of pre-corona and corona days are significantly different. The median of corona days is 95.36% smaller than of pre-corona days with a confidence of 95%. Moreover, the interquartile ranges of pre-corona and corona harvested daily energy are disjoint with a confidence of 95%, as it is visualized in the Figure 6.21. The quantiles of the daily harvested energy are listed in Table 6.7.

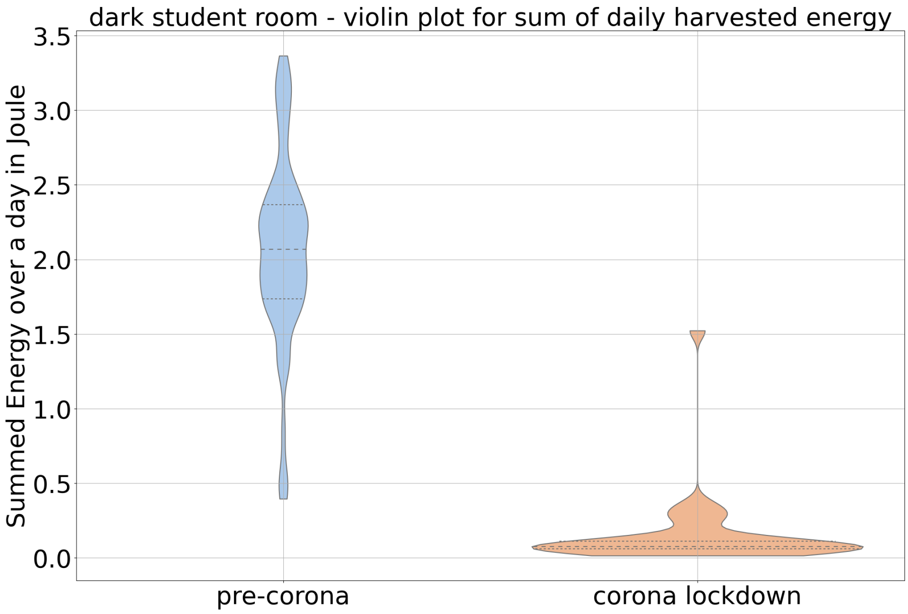


Figure 6.21: Daily based violin plot of the measurement station 13 based on lockdown periods data.

Table 6.7: Quantile values of the daily harvested energy in Joule of position 13 for pre-corona and corona lockdown period weekdays. The lower and upper bounds show an interval, which includes the true quantile value with a confidence of 95%.

Property	**Pre-Corona**	**Corona**
25% quantile lower bound	1.634 J	0.052 J
25% quantile	**1.735 J**	**0.061 J**
25% quantile upper bound	1.901 J	0.07 J
Median lower bound	1.921 J	0.064 J
Median	**2.067 J**	**0.076 J**
Median upper bound	2.215 J	0.089 J
75% quantile lower bound	2.242 J	0.082 J
75% quantile	**2.366 J**	**0.111 J**
75% quantile upper bound	2.526 J	0.286 J
Number of days	60	30

Hourly Based Analysis

The quantiles of the targets pre-corona and corona were examined on a hourly basis. Figure 6.22 shows that from 8am until 7pm significantly more energy is harvested during pre-corona than during corona days regarding the hourly median of the indoor PV harvested energy.

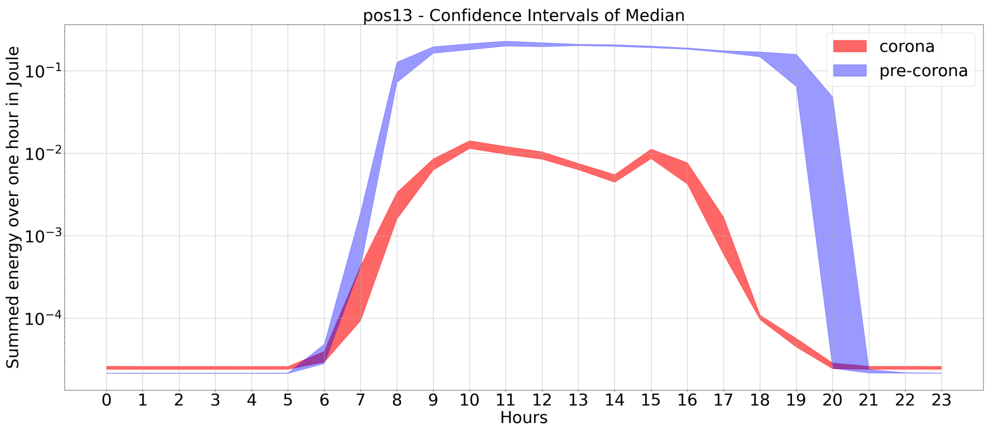


Figure 6.22: Hourly based plot of the measurement station 13 regarding pre-corona (blue) and corona (red) lockdown periods hours. The colorized bands are the 95% confidence intervals of the hourly harvested indoor PV energy median.

6.3.2 Position 14 - Bright Lab

Daily Based Analysis

By examining the data more deeply, it can be concluded that the median of harvested energy at measurement station 14 of pre-corona days is significantly higher ($\alpha = 0.05$) than of corona days. Specifically, the corona median is 51.47% smaller than the pre-corona median with a confidence of 95%. Moreover, as it is depicted in the figure 6.23, the interquartile ranges of pre-corona and corona days are disjoint for this data set only. However, with a confidence of 95% the interquartile range overlaps. The values of the respective quantiles and their bounds are listed in table 6.8.

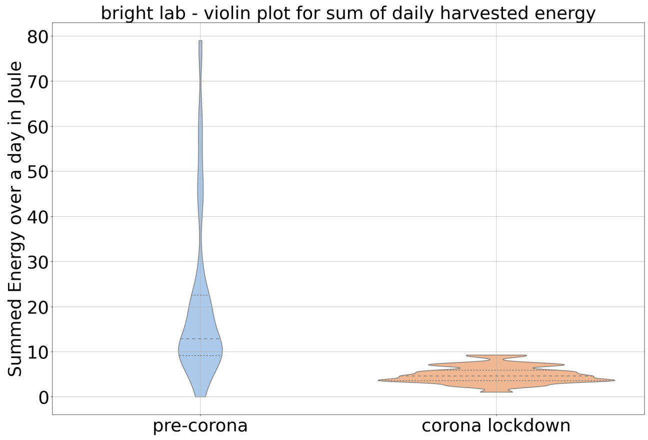


Figure 6.23: Daily based violin lockdown plot of the measurement station 14.

Table 6.8: Quantiles of the daily harvested energy in Joule of position 14 for pre-corona and corona lockdown period weekdays. The lower and upper bounds show an interval, which includes the true quantile with a confidence of 95%.

Property	**Pre-Corona**	**Corona**
25% quantile lower bound	6.93 J	2.68 J
25% quantile	**9.25 J**	**3.58 J**
25% quantile upper bound	10.51 J	4.44 J
Median lower bound	11.23 J	3.81 J
Median	**13.25 J**	**4.58 J**
Median upper bound	18.02 J	5.45 J
75% quantile lower bound	18.73 J	5.00 J
75% quantile	**22.52 J**	**5.88 J**
75% quantile upper bound	43.89 J	7.26 J
Number of days	59	30

Hourly Based Analysis

The quantiles of the targets pre-corona and corona are examined on a hourly basis. Figure 6.24 shows that from 10am until 8pm significantly more energy is harvested during pre-corona than during corona days regarding the hourly median of the indoor PV harvested energy.

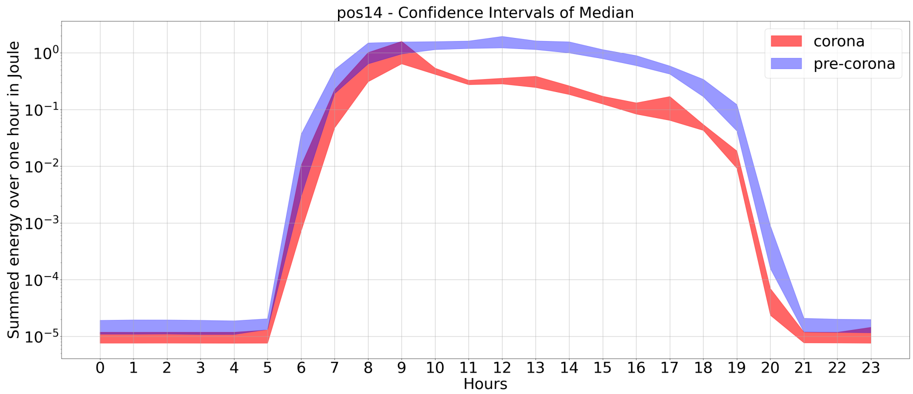


Figure 6.24: Hourly based plot of the measurement station 14 regarding pre-corona (blue) and corona (red) lockdown periods hours. The colorized bands are the 95% confidence intervals of the hourly harvested indoor PV energy median.

6.3.3 Position 17 - Bright Office on Wall

Daily Based Analysis

Figure 6.25 shows that the interquartile ranges of pre-corona and corona days do strongly overlap. Moreover, the pre-corona median of the daily summed harvested energy at measurement station 17 (3.47 J) is even smaller than the corona median (3.57 J). However, these medians do not differ statistically significantly ($\alpha = 0.05$).

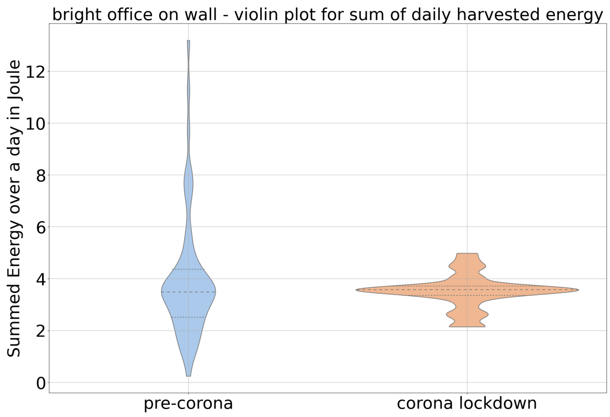


Figure 6.25: Daily based violin lockdown plot of the measurement station 17.

Hourly Based Analysis

Having done the hourly based analysis of the harvested energy for measurement station 17, it must be concluded that not one pre-corona hour 95% confidence interval of the median is disjoint from the corona hour median confidence interval. Figure 6.26 shows the 95% confidence intervals of the hourly summed harvested energy median for pre-corona and corona hours during the whole day.

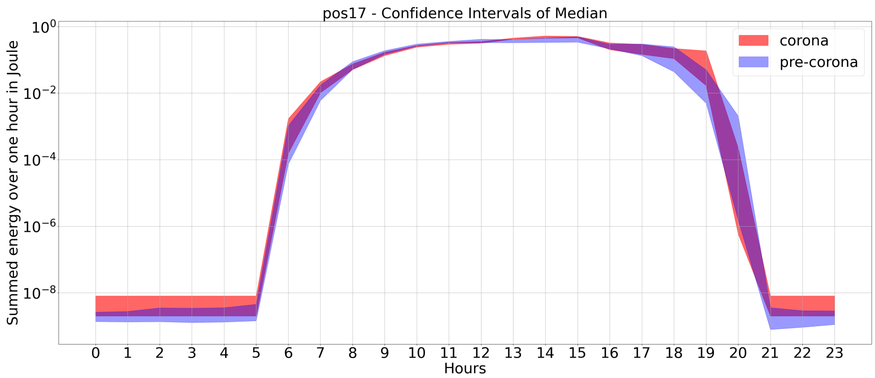


Figure 6.26: Hourly based plot of the measurement station 17 regarding pre-corona (blue) and corona (red) lockdown periods hours. The colorized bands are the 95% confidence intervals of the hourly harvested indoor PV energy median.

6.3.4 Summary of the daily based results

Table 6.9 summarizes all the answers based on the lockdown periods daily based data for each position to the questions:

1. Are the 95% confidence intervals of the pre-corona versus corona median values disjoint? If yes, the relative difference is given. → column *rel. dif. in median* in Table 6.9
2. Is the intersection of the interquartile ranges of the pre-corona versus corona distributions an empty set with a confidence of 95%? → column *Middle 50% disjoint* in Table 6.9

Table 6.9: Summary of the most important statistical findings of the lockdown period daily based analysis for each measurement position.

Position	rel. dif. in median	Middle 50% disjoint
13	-95.36%	✓
14	-51.47%	✗
17	no sign. dif.	✗

6.4 People Detection Period

The indoor harvested energy data trace of a single room (named as *bright office*, see 3.1 for a detailed description of the room properties) is analyzed regarding people presence, which is estimated by the deployed people detection system (see Section 4). The examined time period is about one month. Two measurement stations are within this room. Station 16 lies on a table, which means that the solar cell faces directly the artificial light source. However, station 17 is mounted on a wall, which means that little artificial light reaches directly the corresponding solar cell. Since the stations 16 and 17 are in the same room, the results are very much the same regarding the shape of their distributions of the harvested energy as well as the absolute prediction errors. The biggest difference is seen in the amount of harvested energy due to their mounting specifications. Consequently, the results are only shown for the measurement station 16 and the qualitative results can be easily transferred to the measurement station 17.

In the following the results of the comparison of the different people detection methods are shown (see Section 6.4.1). Then, the results of linking the absolute number of people present inside a room to the harvested energy are given (see Section 6.4.2). After that in Section 6.4.3, the results of the hourly based analysis of the indoor harvested PV energy with respect to room occupancy is shown. Finally, the results regarding the two simple energy prediction schemes are given in Section 6.4.4.

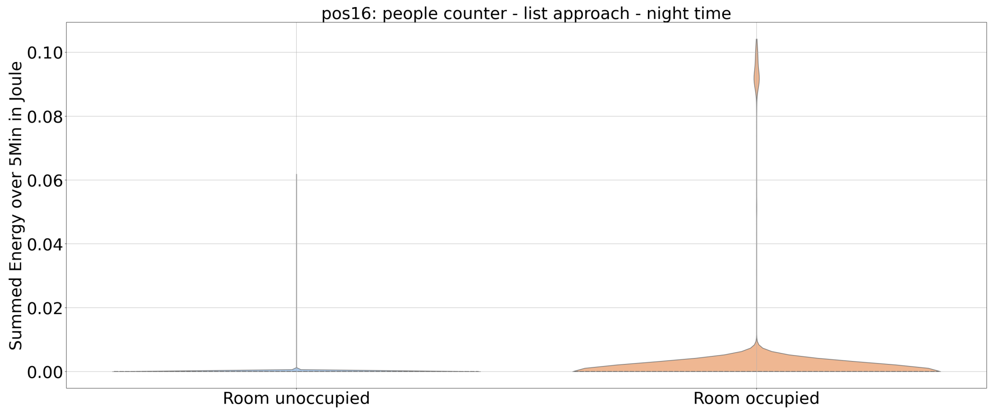
6.4.1 People Detection Methods - Comparison

In addition to the evaluation of the indoor solar harvested energy regarding people presence, the different methods of detecting and counting people are compared as a first step. There are in total seven different methods to estimate people presence. Four of them are binary filters, which are explained in more detail in Section 4.2.1. The other three filters are based on the people counting methods for which an additional binary mapping is done (see Section 4.2.2 for more details). Since there is no data of ground truth available for this period, the seven different methods are compared by looking at the harvested energy at the measurement station 16 during the night times (that is times between sunset and sunrise) over the people detection time period. The idea is that for times, when the room is assumed to be occupied, it holds that the higher the mean value of the harvested energy is, the better is the detection method. The reason for this proceed is the assumption, that people turn always on the lights if they are inside this room at night, which means that people presence is equivalent with switched on lighting. Consequently, people presence is similar to high amount of harvested energy, whereas people absence is similar to low amount of harvested energy. However, not only the mean is considered how good or bad an estimation is, but also the distribution of the underlying data is considered. Besides this, mean and not median values are analyzed here, because of the distribution shape of two strong peaks (see Figures 6.27 and 6.28). Thus, the median could vary a lot between the approaches depending on the fact that it would lie either in the lower or upper peak region. Consequently, mean values are in this case chosen for better comparison, because the mean weights the number of high and low energy values, whereas the median more or less makes a binary choice of either high or low energy.

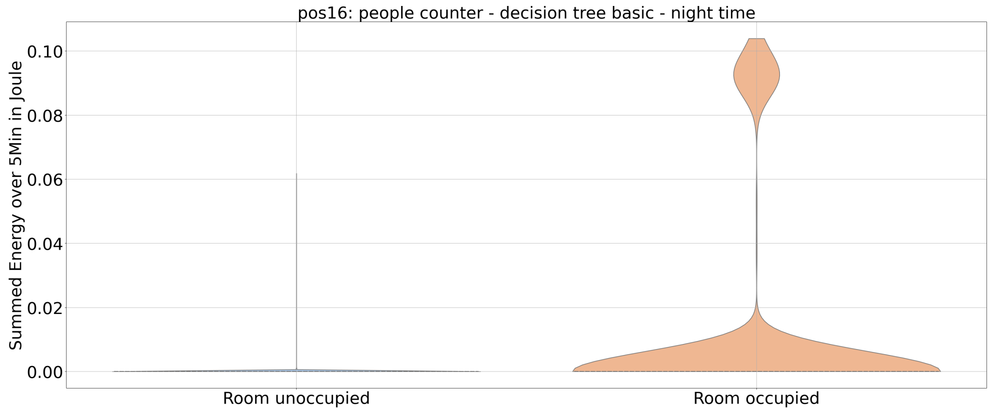
Considering the mean values in table 6.10 and the shapes of the different distributions in Figures 6.27 and 6.28, it is concluded that the best estimation of people presence on this data is made by the *people counter enhanced* method, because it has the highest mean and a high upper peak region in its occupied orange violin (see Figure 6.27c). However, all binary filters perform in terms of the mean almost as good as the *people counter enhanced* method. Nevertheless, the *exponential no sum* method has the highest mean among the binary filter approaches. The mean of the people counter *basic* and *list approach* method is less than 20% of the people counter enhanced mean. Moreover, by looking at the distribution of them in Figures 6.27a and 6.27b more small than high energy harvested values can be seen, which indicates that their estimate of people presence is more wrong than right. Another worth mentioning point is that the *people counter enhanced* method is only working in a retrospective manner and can not be used for a real time estimation whereas all other methods can be used for a real time estimation. Consequently, for a real time estimation, all the binary filter methods which only need one PIR sensor outperform on the analyzed data the implemented people counter methods, which need two PIR sensors to work.

Table 6.10: Mean values of all seven people detection methods of the harvested energy summed over 5 seconds of the measurement station 16 during night times, when the room is estimated to be occupied. Moreover, the column *# of data points* gives the number of 5 seconds intervals, when the room is estimated to be occupied by that detection method over the whole analyzed period of roughly one month.

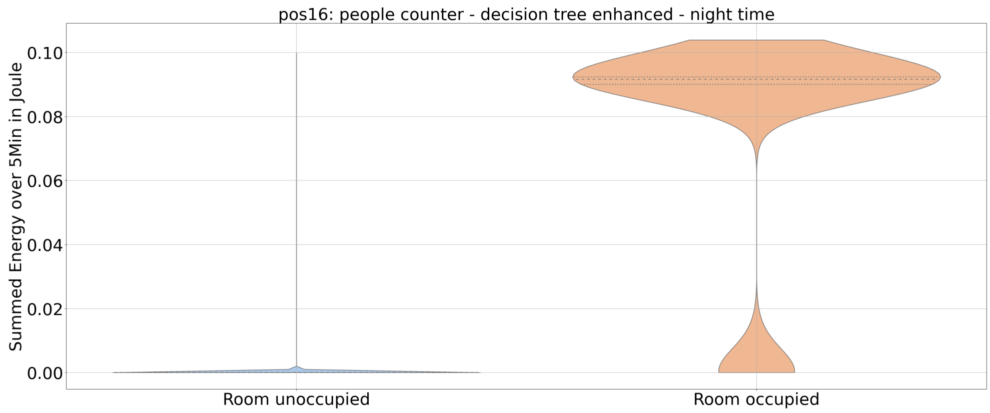
Detection Method	**Mean Value**	**# data points**
Binary Filter - Exp. Simple	$0.894 \cdot 10^{-03}$ J	7162
Binary Filter - Exp. No Sum	$1.207 \cdot 10^{-03}$ J	3689
Binary Filter - Exp. Sum Last 2	$1.022 \cdot 10^{-03}$ J	5803
Binary Filter - Specific	$1.088 \cdot 10^{-03}$ J	4149
People Counter - List approach	$0.036 \cdot 10^{-03}$ J	176338
People Counter - Basic	$0.22 \cdot 10^{-03}$ J	27692
People Counter - Enhanced	$1.314 \cdot 10^{-03}$ J	4240



(a) List approach

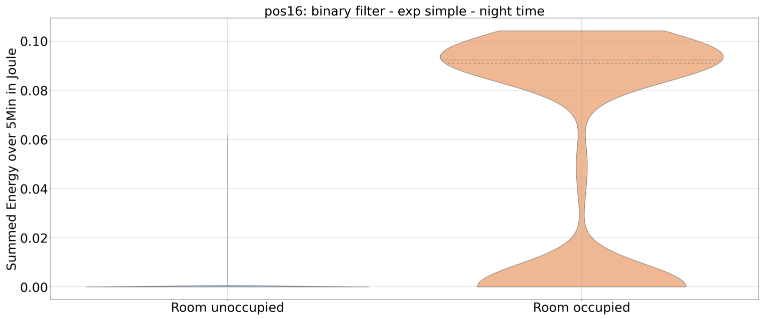


(b) Decision Tree Basic

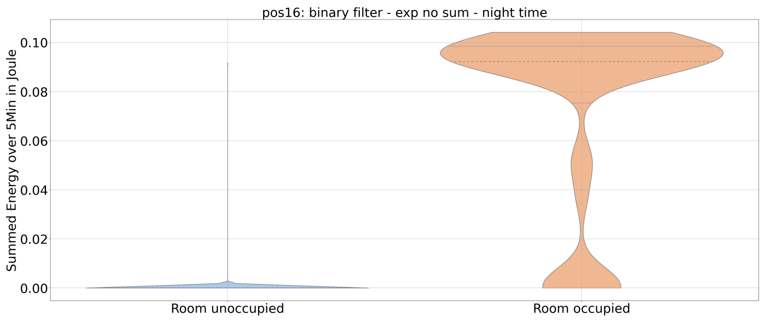


(c) Decision Tree Enhanced

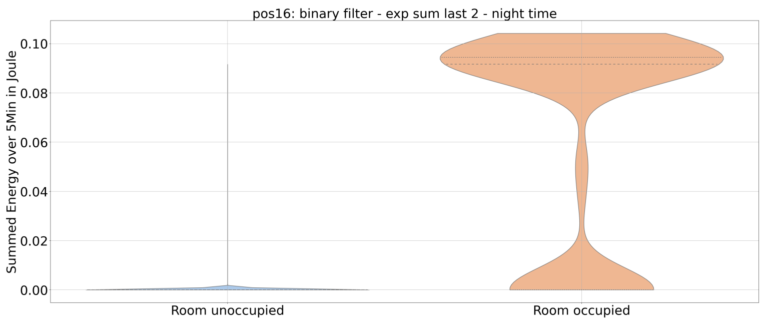
Figure 6.27: Violin plots of the harvested energy summed over 5 minutes of measurement station 16 during night time of the people counting approaches with a binary mapping for people presence estimation. On the left hand side of each plot the blue violin shows energy measurements during estimated unoccupied times. On the right hand side the orange violin shows energy measurements during estimated occupied times.



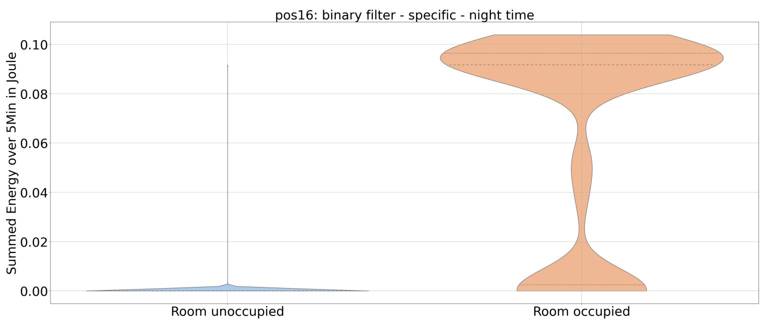
(a) Exp. Simple



(b) Exp. no sum



(c) Exp. sum last 2

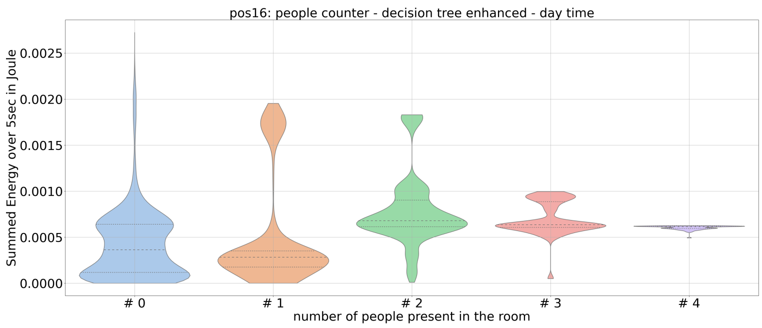


(d) Specific

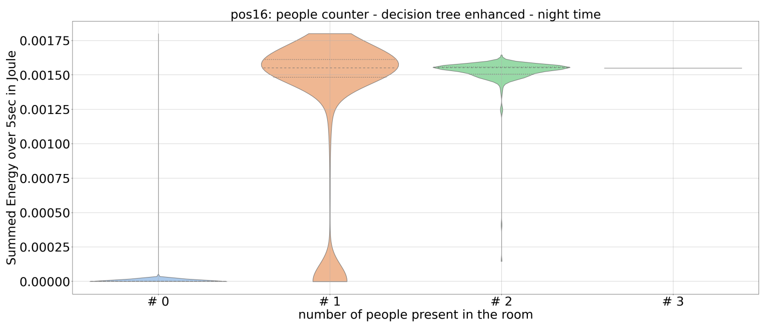
Figure 6.28: Violin plots of the harvested energy summed over 5 minutes of measurement station 16 during night time of the binary filters for people presence estimation. On the left hand side of each plot the blue violin shows energy measurements during estimated unoccupied times. On the right hand side the orange violin shows energy measurements during estimated occupied times.

6.4.2 Absolute number of people present

The analysis regarding the number of people present in a room does not allow to make any statement. The supposedly best people counting method of this work *decision tree enhanced* is shown representative for all the methods in Figure 6.29. As already mentioned in Section 4.3, errors in the people counting application of this work propagate until the next reset event. This means that for a lot of times, the actual number of people present can differ from the estimate. Even though the *decision tree enhanced* method is supposed to be the one with the least error estimates of the three methods (see 6.4.1), the distributions in Figure 6.29 are misleading. For example during night times, the statement that more energy is harvested when two people are in the room compared to one, can not be answered. The Figure 6.29b shows only, that the error probability for a wrong occupancy state estimation is smaller, when the counter is higher. Consequently, all the implemented people counting methods in this work, which uses two PIR sensors for the motion detection have too much faulty estimates as they could be used to make a statement about the causality of the number of people present in the room and the amount of indoor harvested PV energy.



(a) Day time



(b) Night time

Figure 6.29: Violin plots of the harvested energy summed over 5 seconds of measurement station 16 during night and day time of the people counting method *decision tree enhanced*.

6.4.3 Hourly Based Analysis

The harvested energy is summed over 5 seconds, which represents a single data point. When the 5 seconds lie e.g. in between 6am and 7am, then the data point is considered to be in hour 6am. Moreover, all the data points are grouped by hours and the presence of people. That means, for each hour the mean value of the data points, when the room is estimated to be occupied or unoccupied, is calculated. Only the hours are listed in Table 6.11, which had at least one data point for each of the estimations (occupied and unoccupied). In addition to that, only the values with the estimation method *binary filter exponential no sum* are considered, because this is considered to be the most accurate method for real time estimation (see Section 6.4.1). During the hours 6am and 5pm until 8pm the mean value of the harvested energy, when the room is estimated to be occupied, is more than a magnitude higher compared to the unoccupied case (see dark gray rows in Table 6.11). For the hours 7am, 3pm and 4pm the mean value of the harvested energy with people presence is more than twice as high than with people absence (see light gray rows in Table 6.11). However, during the hours 8am until 2pm more energy regarding the mean value is harvested, when no one is estimated to be inside compared to someone is inside. Nevertheless, the difference during these hours in the mean values are less than 100% of the smaller value. Another worth mentioning fact is that the shown data is collected during the end of October and November in Switzerland, which means that the dark gray colorized hours in Table 6.11 describe times after sunset and before sunrise. To get another view of the hourly data, the median values and its confidence intervals are analyzed. In Figure 6.30 the 95% confidence intervals of the median of the harvested energy can be seen. For 6am, 7am and 3pm until 8pm it is significantly more energy harvested in times when the room is estimated to be occupied compared to unoccupied times. However, from 8am till 2pm there is significantly more energy harvested when the room is estimated to be unoccupied compared to occupied.

Table 6.11: Mean values of the harvested energy summed over 5 seconds of the measurement station 16 during the months October and November 2020. The estimates of the room states *occupied* and *unoccupied* is done with the binary filter exponential no sum method. Moreover, the column *# of data points* gives the number of 5 seconds intervals when the room was estimated to be unoccupied/occupied.

Hour	Unoccupied	Occupied	# data points
6	$1.034 \cdot 10^{-06}$ J	$6.309 \cdot 10^{-04}$ J	26574/66
7	$1.676 \cdot 10^{-05}$ J	$4.236 \cdot 10^{-04}$ J	26058/582
8	$1.676 \cdot 10^{-04}$ J	$7.187 \cdot 10^{-05}$ J	26403/237
9	$3.65 \cdot 10^{-04}$ J	$1.028 \cdot 10^{-04}$ J	26517/123
10	$5.223 \cdot 10^{-04}$ J	$3.691 \cdot 10^{-04}$ J	25997/643
11	$5.648 \cdot 10^{-04}$ J	$5.149 \cdot 10^{-04}$ J	25862/778
12	$6.284 \cdot 10^{-04}$ J	$3.341 \cdot 10^{-04}$ J	25847/793
13	$6.126 \cdot 10^{-04}$ J	$3.755 \cdot 10^{-04}$ J	25669/971
14	$5.656 \cdot 10^{-04}$ J	$1.419 \cdot 10^{-04}$ J	25703/937
15	$4.417 \cdot 10^{-04}$ J	$1.154 \cdot 10^{-03}$ J	25697/943
16	$1.826 \cdot 10^{-04}$ J	$9.138 \cdot 10^{-04}$ J	25393/1247
17	$3.918 \cdot 10^{-05}$ J	$1.418 \cdot 10^{-03}$ J	24993/1647
18	$1.955 \cdot 10^{-05}$ J	$1.188 \cdot 10^{-03}$ J	25751/889
19	$2.588 \cdot 10^{-05}$ J	$1.093 \cdot 10^{-03}$ J	26186/454
20	$3.953 \cdot 10^{-06}$ J	$6.366 \cdot 10^{-04}$ J	26432/208

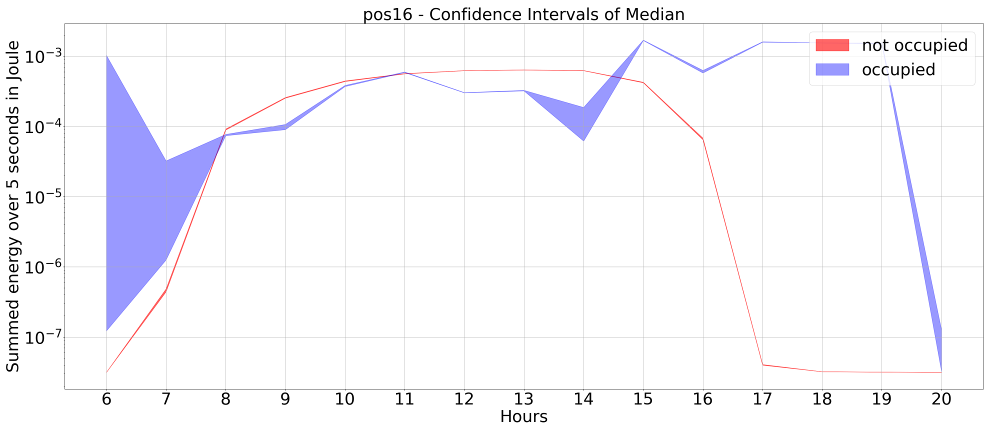


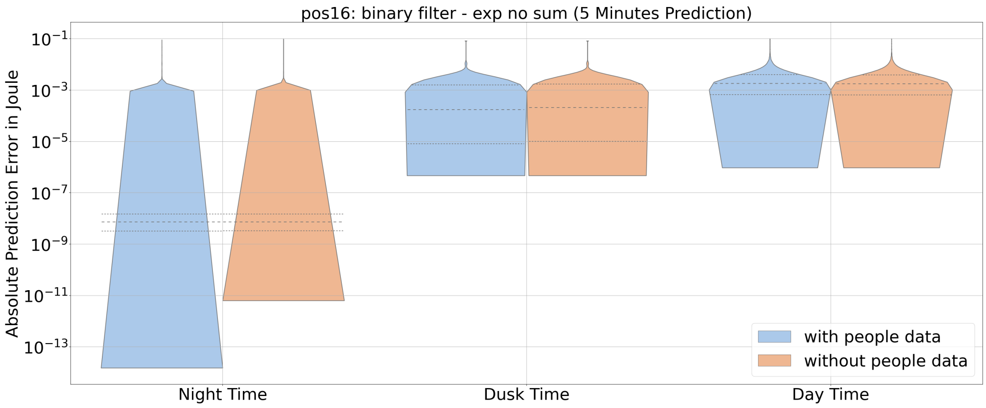
Figure 6.30: Hourly plot of the harvested energy summed over 5 seconds of measurement station 16 with the occupancy detection method *binary filter exponential no sum*. The blue area shows the 95% confidence interval of the over 5 seconds summed in a specific hour harvested energy median of the occupied times, whereas the red area the one of the occupied times.

6.4.4 EWMA - Prediction Error Analysis

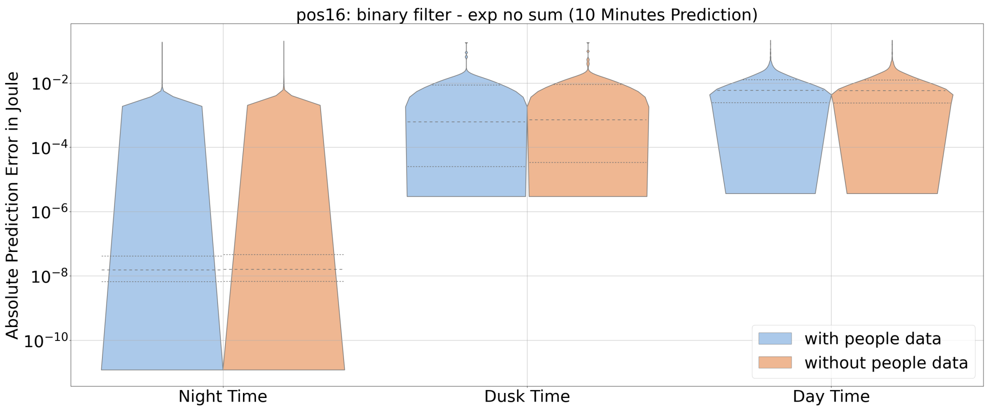
The augmented EWMA prediction is done for all seven people detection approaches. However, only the results of the *binary filter exponential no sum* is shown here, as it is according to section 6.4.1 the best performing method with a potential real time implementation. The absolute prediction error is compared between a EWMA prediction which takes people detection data into account (blue in all three plots) and another one which does not (orange in all three plots). They are compared during night, dusk and day times and for different prediction horizons (5 minutes, 10 minutes and 1 hour). The distributions of the absolute prediction errors are shown with a log scale on the y-axis in all three Figures 6.31a, 6.31b and 6.31c. The distributions of the absolute prediction errors with time horizons 5 and 10 minutes for night, dusk and day time regarding taking people detection data into account or not are very similar in shape. This means that their medians do not deviate significantly from each other and their interquartile ranges do greatly overlap (see Figures 6.31a and 6.31b). The same is true for the prediction horizon of one hour during dusk and day times (see Figure 6.31c). However, the medians of the absolute errors during night time differ statistically significantly ($\alpha = 0.05$) and the median with taking people detection into account is 10.91% less of the median where no people detection data is considered with a confidence of 95%. The corresponding values are listed in table 6.12.

Table 6.12: Median values of absolute prediction error on the hourly harvested energy in Joule of position 16 for one EWMA prediction scheme which takes people presence data into account and another which does not. The lower and upper bounds show an interval, which includes the true median value with a confidence of 95%.

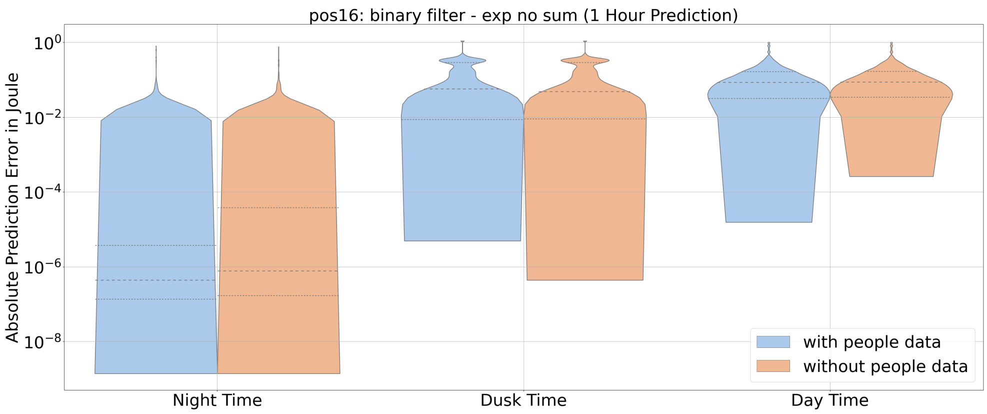
Property	**With people data**	**Without people data**
50% Median lower bound	$3.816 \cdot 10^{-07}$ J	$6.0 \cdot 10^{-07}$ J
50% Median	$4.378 \cdot 10^{-07}$ J	$7.654 \cdot 10^{-07}$ J
50% Median upper bound	$5.345 \cdot 10^{-07}$ J	$1.005 \cdot 10^{-06}$ J
Number of hours	499	499



(a) Prediction horizon of five minutes



(b) Prediction horizon of ten minutes minutes



(c) Prediction horizon of one hour minutes

Figure 6.31: Violin plot of absolute prediction error of the two prediction schemes at measurement station 16 with the people detection binary filter method *exponential no sum* with different prediction horizons.

Chapter 7

Discussion

7.1 Measurement Station Result Statements

Hourly and daily based analysis of six different indoor harvested energy measurement stations are conducted. Each station has its own light characteristics. That is why in a first round, the important results of each station are discussed separately. Afterwards a more general way of formulating the findings is presented.

Position 06 - Dark office
: Since there is no lockdown data for this position, the best achieved ground truth approximation for room occupancy in this work is the differentiation between weekdays (people presence) and weekends (people absence). However, since this room is a single person office and not a multi person office or a student room, the probability that the room is occupied during weekends is reasonably low. The results show that significantly more energy is harvested during weekdays than during weekends. Consequently, the statement is allowed that people presence lets the indoor harvested PV energy increase on a daily and hourly basis at this position. Besides this, the seasonal influence on the indoor harvested PV energy at this station is only marginal. Finally, it is worth mentioning that this measurement station has only little natural light.

Position 13 - Dark Student Room
: Since there is lockdown data available for this station, the comparison between the weekdays (people presence) and weekends (people absence) can be seen as a first hint, how people presence influence the indoor harvested PV energy at this station. Because it is no secret that students also work on weekends, this distinction between weekdays and weekends is not best suited to conclude something about people presence (weekdays) or absence (weekends). Nevertheless, the result show that more energy is harvested during weekdays compared to weekends. However, looking at a better

ground truth approximation with pre-corona (people presence) and corona (people absence) lockdown periods, the result show that significantly more energy is harvested during pre-corona compared to corona lockdown periods. Having this better ground truth approximation, it is concluded that people absence let the indoor harvested PV energy significantly shrink at this position. Since there is only some natural light at the measurement station, the seasonal differences do not lead to similarly strong effects as it is the case for the people presence.

Position 14 - Bright Lab

The results of the whole years analysis for this measurement station show that the seasonal differences have significant impact on the indoor harvested PV energy. Most energy is harvested during summer, then spring, fall and winter. With the weekdays (people presence) and weekends (people absence) ground truth approximation of room occupancy, no significant difference in the harvested energy is observed. However, with a better ground truth approximation on room occupancy during lockdown periods analysis, the results show that even during daylight hours little more energy is harvested during pre-corona compared to corona lockdown periods. Nevertheless, because there is significant natural light at this measurement station, room occupancy is not the main indicator on how much energy is harvested. Both, room occupancy as well as the current season affect the indoor harvested PV energy to a similar extent.

Position 16 - Bright Office on Table

There is no whole years and no lockdown periods analysis for this measurement station. Only the analysis with the people detection system is done. The qualitative findings during this time are the same as for position 17, which is located in the same room as position 16. The findings are described in more detail in the next discussion paragraph of position 17.

Position 17 - Bright Office on Wall

The results of the whole years analysis for this measurement station show that the seasonal differences have significant impact on the indoor harvested PV energy. Most energy is harvested during summer, then spring, fall and winter. With the weekdays (people presence) and weekends (people absence) ground truth approximation of room occupancy, no significant difference in the harvested energy is observed. Even with a better ground truth approximation of room occupancy during lockdown periods analysis, the results show that at no hour during the day, the indoor harvested PV energy is significantly influenced by the room occupancy. Going a step further and having the low power people detection system installed at the door of this room, the harvested energy during people presence is only significantly higher when it is dark outside. During day times when it is bright outside, even more energy is harvested, when people are absent. This strange behaviour during daylight hours is most likely caused by weather differences. There was not enough data gathered with the people detection system such that weather fluctuations cancel out. Nevertheless, because there is significant natural light at this measurement station, room occupancy is not the main indicator on how much energy is harvested. The impact of the different seasons is greater than the impact of the room occupancy on the amount of indoor harvested PV energy for this measurement station during daylight times.

Position 18 - Hallway
Since there is no lockdown data for this position, the best achieved ground truth approximation for room occupancy in this work at this position is the distinction between weekdays and weekends. However, since the hallway is not a single person office, the probability, that no person is in the hallway during whole weekends, is reasonably not very high. This means that this ground truth approximation of people presence (weekdays) and absence (weekends) might not be very good. Nevertheless, a tendency is visible based on the distributions and the numerical values of the data that people presence might lead to an increase in the indoor harvested PV energy in the hallway. To have a stronger statement for this measurement station, data with a better ground truth approximation of people presence is needed. Besides this, the seasonal impact on the amount of harvested energy is marginal, which is justified by having no natural light at the hallway measurement station.

Even though the station characteristics are different at each station, a distinction into two types of positions is made to summarize the findings in a more general manner. One group of the measurement stations is artificial light driven, which includes the stations 06, 13 and 18. The other group is natural light driven, which includes the stations 14, 16 and 17. However, this classification makes only sense during times when it is bright outside, which is between sunrise and sunset. Because, when it is between sunset and sunrise, i.e. dark outside, every measurement station is artificial light driven. Nevertheless, it is concluded that for the artificial light driven stations people presence lead to a significant increase in the amount of indoor harvested PV energy. However, the seasonal impact on the harvested energy is marginal for these stations. On the other side, it is concluded that for natural light driven stations people presence might or might not lead to an increase in the amount of harvested energy depending on how big the influence of the artificial light still is. The bigger the influence of the artificial light compared to the natural light, the more impact has people presence on the harvested energy. However, for the natural light driven measurement stations it is very likely, that seasonal differences have an effect on the indoor harvested PV energy.

7.2 People Detection System

Having implemented seven different people detection methods of which four uses only one PIR sensor and three uses two PIR sensors, it is concluded that it is enough to detect room occupancy with only one PIR sensor. This conclusion is based on the collected data and the investigated methods, which are by no means perfect. Especially the people counter algorithms, which are suitable for real time applications, could be enhanced. Moreover, only one specific room was analyzed. In another room with other characteristics (more people, more space, etc.) the findings might be not the same. Nevertheless, a tendency can be observed that low power solutions with one PIR sensor are accurate enough to detect room occupancy for connecting people presence to the indoor harvested PV energy.

7.3 Energy Prediction

Two simple prediction schemes for three different prediction horizons are evaluated at the measurement station 16 and 17. Comparing the scheme, which makes use of the room occupancy information, to the other scheme, which does not consider this information, it is concluded that there is no performance difference between them during night, dusk and day times with time horizons of five and ten minutes regarding the absolute prediction errors. The same can be stated for the one hour prediction horizon during dusk and day time. However, during night time and a prediction horizon of one hour, the prediction with considering people detection data performs better than the other. Moreover, the whole years and lockdown period analysis does not reveal any influence of the people presence on the harvested energy at station 17. Meaning, for both stations 16 and 17, which are in the same room, the artificial light contributes very little energy compared to the natural light during the daylight hours. This means that people presence might not drastically impact the amount of indoor harvested PV energy in this case. Consequently, even the best and most accurate prediction algorithm might have problems to perform better with the people detection data than without. Nevertheless, the night time is interesting, because during these times the natural light contributes very little to the harvested energy. This means that the night time results of the prediction of this room might be transferred to a more general artificial light driven position. Hence, based on these results and the room characteristics, a hypothesis can be formulated. People presence data improves the prediction on indoor harvested PV energy in artificial light driven positions with a prediction horizon of one hour or bigger. However, this hypothesis is only indicated and by far not proven with the results of this work. More people detection data needs to be collected in various positions such that this hypothesis can be confirmed or refuted.

Chapter 8

Conclusion and Outlook

Since there was no ground truth data available about the room occupancy, several approximations of the ground truth were implemented. One of them was to use a low power people detection system, which used PIR sensors to detect people. Within that system, seven approaches to estimate the room occupancy were examined. Based on the collected data and the observed room characteristic, the people detection solution with only one PIR sensor outperformed the people counting methods with two PIR sensors.

Six indoor harvesting PV energy measurement stations each with their own lighting conditions were analyzed with respect to the indoor harvested PV energy in order to answer the following question. How does people presence influence the amount of indoor harvested PV energy? Even though only approximations of the people presence ground truth were used, the following can be concluded. It is found that in settings, where most of the available light is originated from artificial light sources, people presence leads to an increase in the harvested energy. In settings where most of the available light is originated from natural light, people presence has a minor or no influence. In those cases seasons (spring, summer, fall or winter) have a stronger impact than people presence on the amount of indoor harvested PV energy.

In order to also predict the future availability of PV energy, the collected people detection data was combined with the existing indoor harvested PV energy measurement data. The evaluation of the two implemented augmented EWMA prediction schemes show signs that not only the harvested data depends on people presence, but also prediction of future harvested energy can be improved at artificial light driven places with the knowledge of room occupancy. However, dealing with energy prediction was only a minor part in this work. More rooms need to be complemented with a people detection system and more data over a longer time period is needed to implement better prediction algorithms.

In future work it would be interesting to have low power people detection systems complemented with ground truth data about the room occupancy. With this, the performance of the low power system could be measured. Moreover, complementing each room, where at least one indoor energy harvesting measurement station is placed, with a people detection system would help to confirm or refute the statements made in this work. Besides this, a possible next step could be to have a low power people detection system, which itself can harvest and measure indoor PV energy. Thus, this would mean to go one step closer to the system design of a low power sensor node, which has an on-board PV harvesting solution and uses people detection data to improve its performance.

Bibliography

[1] Thomas Huld, Richard Müller, and Attilio Gambardella. A new solar radiation database for estimating pv performance in europe and africa. *Solar Energy*, 86(6):1803 – 1815, 2012. ISSN 0038-092X. doi: https://doi.org/10.1016/j.solener.2012.03.006. URL `http://www.sciencedirect.com/science/article/pii/S0038092X12001119`.

[2] Lukas Sigrist, Andres Gomez, and Lothar Thiele. Dataset: Tracing indoor solar harvesting. In *Proceedings of the 2nd Workshop on Data Acquisition To Analysis*, DATA'19, page 47–50, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450369930. doi: 10.1145/3359427.3361910. URL `https://doi.org/10.1145/3359427.3361910`.

[3] Marvin Häberle. People counting system using wireless sensor nodes. Technical Report 2, ETH Zürich, 4 2020. By Prof. Dr. Lothar Thiele with supervisors: Andreas Biri and Stefan Drašković.

[4] Clément Antoine Reynaud, Raphael Clerc, Pierre Balthazar Lechêne, Mathieu Hébert, Anthony Cazier, and Ana Claudia Arias. Evaluation of indoor photovoltaic power production under directional and diffuse lighting conditions. *Solar Energy Materials and Solar Cells*, 200:110010, 2019. ISSN 0927-0248. doi: https://doi.org/10.1016/j.solmat.2019.110010. URL `http://www.sciencedirect.com/science/article/pii/S0927024819303393`.

[5] X. Ma, S. Bader, and B. Oelmann. Power estimation for indoor light energy harvesting systems. *IEEE Transactions on Instrumentation and Measurement*, 69(10):7513–7521, 2020. doi: 10.1109/TIM.2020.2984145.

[6] Andres Gomez, Francesco Conti, and Luca Benini. Thermal image-based cnn's for ultra-low power people recognition. pages 326–331, 05 2018. doi: 10.1145/3203217.3204465.

[7] A. R. Pratama, A. Lazovik, and M. Aiello. Office multi-occupancy detection using ble beacons and power meters. In *2019 IEEE 10th Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*, pages 0440–0448, 2019. doi: 10.1109/UEMCON47517.2019.8993008.

[8] F. Wahl, M. Milenkovic, and O. Amft. A distributed pir-based approach for estimating people count in office environments. In *2012 IEEE 15th International Conference on Computational Science and Engineering*, pages 640–647, 2012. doi: 10.1109/ICCSE.2012.92.

[9] S. H. Kim, H. J. Moon, and Young Ran Yoon. Improved occupancy detection accuracy using pir and door sensors for a smart thermostat. 2017.

[10] Sidharth Mishra, Uttam Sarkar, Subhash Taraphder, Sanjoy Datta, Devi Swain, Reshma Saikhom, Sasmita Panda, and Menalsh Laishram. Principal component analysis. *International Journal of Livestock Research*, page 1, 01 2017. doi: 10.5455/ijlr.20170415115235.

[11] Seaborn Documentation. Violin plot — seaborn documentation, 2021. URL `https://seaborn.pydata.org/generated/seaborn.violinplot.html`. [Online; accessed 08-January-2021].

[12] William R. Thompson. On confidence ranges for the median and other expectation distributions for populations of unknown distribution form. *Ann. Math. Statist.*, 7(3):122–128, 09 1936. doi: 10.1214/aoms/1177732502. URL `https://doi.org/10.1214/aoms/1177732502`.

[13] Aman Kansal, Jason Hsu, Sadaf Zahedi, and Mani B. Srivastava. Power management in energy harvesting sensor networks. *ACM Trans. on Embedded Computing Sys.*, 6:32, December 2007. URL `https://www.microsoft.com/en-us/research/publication/power-management-in-energy-harvesting-sensor-networks/`.

[14] Schweizer Tourismus Verband. Coronavirus: Chronik mit bezug zum schweizer tourismus, 2021. URL `https://www.stv-fst.ch/de/chronik-coronavirus`. [Online; accessed 08-January-2021].

Appendix A

Appendix

A.1 People Counting - Basic Decision Tree Approach

Algorithm 1 Basic Decision Tree Approach

Input: event_list with three types of events (midnight_reset, motion_inside and motion_outside)
Output: List of people counter and corresponding timestamp.

```
people_counter_event_list = empty_list
pc = 0
last = 0
next = 1
while last < length(event_list) do
  if event_list[last].type() == midnight_reset then
    pc = 0
    people_counter_event_list.append(pc, event_list[last].timestamp())
    last += 1, next += 1 continue
  else if event_list[last].type() == motion_inside then
    if event_list[next].type() == motion_outside and timestamps of last and
    next are in range then
      pc = max(pc - 1, 0)
      people_counter_event_list.append(pc, event_list[last].timestamp())
      last += 2, next += 2, continue
    else if pc == 0 then
      pc += 1
      people_counter_event_list.append(pc, event_list[last].timestamp())
      last += 1, next += 1, continue
    end if
  else if event_list[last].type() == motion_outside then
    if event_list[next].type() == motion_inside and timestamps of last and
    next are in range then
```

```
        pc += 1
        people_counter_event_list.append(pc, event_list[last].timestamp())
        last += 2, next += 2, continue
      end if
    end if
    last += 1, next += 1
end while
return people_counter_event_list
```

A.2 People Counting - Enhanced Decision Tree Approach

Algorithm 2 Enhanced Decision Tree Approach

Input: event_list with four types of events (midnight_reset, motion_inside, motion_outside and last_inside_motion_reset)
Output: List of people counter and corresponding timestamp.

```
people_counter_event_list = empty_list
pc = 0
last = 0
next = 1
while last < length(event_list) do
  if event_list[last].type() == midnight_reset then
    pc = 0
    people_counter_event_list.append(pc, event_list[last].timestamp())
    last += 1, next += 1, continue
  else if event_list[last].type() == last_inside_motion_reset then
    pc = 0
    people_counter_event_list.append(pc, event_list[last].timestamp())
    last += 1, next += 1, continue
  else if event_list[last].type() == motion_inside then
    if event_list[next].type() == motion_outside and timestamps of last and
    next are in range then
      pc = max(pc - 1, 0)
      people_counter_event_list.append(pc, event_list[last].timestamp())
      if event_list[next + 1].type() == motion_inside and timestamps of
      next and next + 1 are in range then
        last += 3, next += 3, continue
      end if
      last += 2, next += 2, continue
    else if pc == 0 then
      pc += 1
      people_counter_event_list.append(pc, event_list[last].timestamp())
      last += 1, next += 1, continue
    end if
  else if event_list[last].type() == motion_outside then
    if event_list[next].type() == motion_inside and timestamps of last and
    next are in range then
      pc += 1
      people_counter_event_list.append(pc, event_list[last].timestamp())
      if event_list[next + 1].type() == motion_outside and timestamps of
      next and next + 1 are in range then
```

```
            last += 3, next += 3, continue
        end if
        last += 2, next += 2, continue
      end if
  end if
  last += 1, next += 1
end while
return people_counter_event_list
```