MA 414

E.ON Energy Research Center | RWTH AACHEN UNIVERSITY

Master's Thesis

Evaluation of the use of typical days for approximate optimal control applications in heat pump systems

Design von Masterarbeiten

Aachen, December 2021

Antonio Gil Toral
Matriculation number: 428155

Supervisors:
Laura Maier, M.Sc.
Univ.-Prof. Dr.-Ing. Dirk Müller

The present work was submitted to the:
E.ON Energy Research Center | ERC
Institute for Energy Efficient Buildings and Indoor Climate | EBC
Mathieustrasse 10, 52074 Aachen, Germany

Abstract

Buildings contribute to roughly 40 % of the global energy use, of which approximately 40 % is used for heating, cooling, ventilation and air-conditioning (HVAC). In this context, heat pump (HP) systems are of great importance as they can satisfy the cooling and heating demands of a building with operation efficiencies up to 600 %. According to the current state of research, optimal control strategies applied to HVAC systems can obtain potential energy savings between 13 and 28 %. In this context, model predictive control (MPC) has emerged as an important control strategy on intelligent building operation. Nevertheless, MPC is not widely adopted in practice. As an alternative, an approximate or offline MPC is used. This approach is based on the optimization results of a mixed-integer linear program (MILP), followed by a rule mining process, where approximated control rules for the system are extracted by means of Machine Learning (ML) algorithms.

The use case of this project is the heat pump system of a former military hospital in Berlin, which is being transformed into a business center called FUBIC. The system is modeled and used within an approximate MPC process. In some cases, it is too computationally expensive to perform this optimization system for such large time horizons. This is the reason for the use of typical days to reduce the original inputs (energetic demand and outside temperature) into a defined number of representative days. For the use case of this thesis, this is performed by means of a k-medoid clustering algorithm. Anyhow, it remains to be proven if the final results are accurate enough in order to obtain precise control rules for the system.

The method to evaluate the effects of use of typical days in the approximate MPC process is divided into two different parts: an a priori and an a posteriori analysis. In the first, the performance of the time series aggregation regarding the representation of the input curves will be evaluated using different regression and correlation metrics. In the a posteriori analysis, the outputs of the optimization program will be compared when using the clustering approach for different number of representative days. The last part of the analysis will consist in evaluating the effects of the use of typical days in the rule mining stage of the approximate MPC process. In order to do so, the outputs of the optimization program, when using different number of representative days, are used as training data for the ML tool, which predicts the optimal operation schedule of the HP for a set of unseen data (testing period). This prediction is compared to the optimal operation schedule obtained by running the optimization program with the full year inputs for the same set of testing data. This way, the effects of the use of typical days in different stages of the approximate MPC process are evaluated.

Zusammenfassung

Gebäude tragen zu etwa 40 % des weltweiten Energieverbrauchs bei, von denen etwa 40 % für Heizung, Kühlung, Lüftung und Klimatisierung (HLK) verwendet werden. In diesem Zusammenhang sind Wärmepumpensysteme von großer Bedeutung, da sie den Kühl- und Heizbedarf eines Gebäudes mit Wirkungsgraden von bis zu 600 % decken können. In diesem Zusammenhang hat sich die Model Preditive Control (MPC) als eine wichtige Regelungsstrategie für den intelligenten Gebäudebetrieb erwiesen. Dennoch wird MPC in der Praxis nicht häufig eingesetzt. Als Alternative wird eine approximative oder Offline-MPC verwendet. Dieser Ansatz basiert auf den Optimierungsergebnissen eines Mixed Integer Linear Program (MILP), gefolgt von einem Rule-Mining-Prozess, bei dem mittels Machine Learning (ML) approximierte Steuerungsregeln für das System extrahiert werden.

Der Anwendungsfall in diesem Projekt ist die Wärmepumpenanlage eines ehemaligen Militärkrankenhauses in Berlin, das in ein Geschäftszentrum namens FUBIC umgewandelt wird. Das System wird modelliert und im Rahmen eines approximativen MPC-Prozesses verwendet. In einigen Fällen ist es zu rechenaufwendig. In manchen Fällen ist es zu rechenintensiv, dieses Optimierungsysteme für so große Zeithorizonte durchzuführen. . Dies ist der Grund für die Verwendung typischer Tage, um die ursprünglichen Eingaben (Energiebedarf und Außentemperatur) auf eine bestimmte Anzahl von repräsentativen Tagen zu reduzieren. Für den Anwendungsfall dieser Arbeit erfolgt dies mit Hilfe eines k-medoid Clustering Algorithmus. Es bleibt jedoch zu prüfen, ob die Ergebnisse genau genug sind, um präzise Steuerungsregeln für das System zu erhalten.

Die Methode zur Bewertung der Auswirkungen der Verwendung typischer Tage im approximativen MPC-Prozessgliedert sich in zwei verschiedene Teile: eine a priori und eine a posteriori Analyse. Im ersten, wird die Leistung der Zeitreihenaggregation hinsichtlich der Darstellung der Eingangs Kurven. In der a posteriori Analyse werden die Ergebnisse des Optimierungsprogramms bei Verwendung des Clustering Ansatzes für eine unterschiedliche Anzahl repräsentativer Tage verglichen. Der letzte Teil der Analyse wird darin bestehenmdie Auswirkungen der Verwendung typischer Tage in der Phase der Regelfindung des approximativen MPC-Prozesses. Zu diesem Zweck werden die Ergebnisse des Optimierungsprogramms bei Verwendung einer unterschiedlichen Anzahl von repräsentativen Tagen als Trainingsdaten für das ML-Tool verwendet, das den optimalen Betriebszeitplan der HP für einen Satz unbekannter Daten (Testzeitraum) vorhersagt. Diese Vorhersage wird mit dem optimalen Betriebsplan verglichen, der durch Ausführen des Optimierungsprogramms mit den Eingaben für das ganze Jahr für denselben Satz von Testdaten erhalten wurde. Auf diese Weise

können die Auswirkungen der Verwendung von typischer Tage in den verschiedenen Phasen des approximativen MPC-Prozesses bewertet werden.

Table of Contents

List of Figures

List of Tables

1 Introduction

The latest reports in regard to climate change have made several countries in the last years declare the state of climatic emergency [2]. In this context, the United Nations Climate Change Conference will hold its 26^{th} edition in November 2021 in order to define the next steps in the fight against climate change. Several authorities and members of the scientific community stated that the measures agreed in the previous edition of the conference were not sufficient and that it is necessary to adopt stricter policies in the coming edition. [15]

In view of this situation together with the progressive increase of global energetic demand, the importance of developing efficient energy systems is growing. According to the European Commission, buildings are responsible for approximately 40 % of EU energy consumption and 36 % of the greenhouse gas emissions [6]. Therefore, buildings are the largest energy consumers in Europe, which implies that there is a great potential for optimization in this sector. A large portion of the energy consumed in buildings is used in heating, cooling, ventilation and air conditioning (HVAC). In this context, the integration of heat pump (HP) systems becomes of great importance. These systems can operate with efficiencies between 300 and 500 % [4], and use low temperature sources to extract heat. HPs are electricity based, which allows the use of energy obtained from renewable sources, holding a great decarbonization potential. However, these systems require appropriate control strategies in order to maximize their efficiency. According to the current state of research, optimal control strategies applied to HVAC systems can obtain average energy savings between 13 and 28 %. [17]

HP systems offer great advantages in terms of efficiency and reduction of CO_2 emissions, but their implementation in building energy systems (BESs) also brings challenges. The efficiency of HPs highly depends on the boundary conditions, i.e, source and sink temperature, which causes the need to integrate these technologies in complex systems to exploit their potential. Additionally, the connection to electricity grids fed by renewable sources of energy adds complexity to the system, due to the fluctuations in availability related to renewable energy technologies. Ultimately, this leads to the fact that conventional, rule-based strategies are reaching their limits and the need for modern control methods is growing. One of the modern methods is Model Predictive Control (MPC), which, in the last two decades, has become a dominant control strategy in research on intelligent building operation [17]. The core of such control is the process model, which is used to predict future states while considering a series of constraints. However, MPC is not widely adopted in practice, due to a

number of reasons. The first is that, unlike other fields where MPC have been applied successfully, such as the process industry, building management systems engineers do not have deep training in advanced control strategies [26]. The application of MPC also requires precise mathematical models of the controlled building, which are complex to design. Additionally, the high software and hardware requirements of MPC have also become a barrier to the its wide implementation in practice. [14]. In the recent years, some studies [13, 18, 23] have developed promising solutions to this problem in many different applications by introducing Machine Learning (ML) concepts. The idea is based on extracting simplified control rules from an MPC based optimization that can easily be implemented in the system by means of ML algorithms. These extracted rules have the same format as conventional rule-based approaches, which are easier to interpret and do not require training in advanced control strategies. This concept receives the name of approximate or offline MPC. Although this approach has already been applied in the context of BESs, HP systems have not been assessed yet, even though their application in energy optimization has great potential.

The use case for this thesis is defined in [28]. It consists in an approximate MPC process designed for the HP system that will cover the HVAC demands of FUBIC, a former military hospital in Berlin that is currently being transformed into a business center. The approximate MPC is based on the optimization results of a mixed-integer linear program (MILP), which computes the system's optimal operation in order to minimize its cost. Then, the optimization results go through a rule mining process, where approximated control rules for the system are extracted by means of a Random Forest (RF) classifier. This way, these control rules can be introduced in the HP system for it to learn to operate optimally under different scenarios.

In some cases, it is not possible to have access to a series of inputs for a large time horizon (months or even years), or it is too computationally expensive to optimize the operation of an energy system for these time horizons. In this context, the use of typical or representative days appears as a promising solution for this problem. These typical days try to represent the values, trends and behaviour of larger data sets in a compressed format. This approach has already been proven effective in BESs, leading to reasonable results in representing input curves for this kind of application [16]. However, it remains to be proven if the use of representative days is a promising approach in order to extract generalizable control rules for a HP system in an approximate MPC process. In order to evaluate the effects of the use of typical days in the different stages of the approximate MPC process, the following outputs will be analyzed, when using different number of representative days:

- Inputs after the clustering process
- Operation optimization outputs
- Extracted control rules

2 Theoretical Basics and State of the art

This chapter serves as a theoretical basis to different concepts that will be assessed on this thesis. The main notions that are introduced are data clustering and its use in time series aggregation for energy optimization applications.

2.1 Data clustering

Clustering is the process that creates clusters or groups of data in a way that the objects which fall in a same group share, in some way, similar properties and differ from those which are sorted in other different groups during the process. During the years, clustering has been used in different fields of practice, including Big Data, marketing or even medical applications[11]. Diverse clustering algorithms exist nowadays, although, in this thesis, we will focus on partitioning-based clustering, as the algorithm used in the approximate MPC process will be of this kind.

2.1.1 Partitioning-based clustering

Partitioning or centroid based clustering is a method in which each cluster is represented by a central vector, and the different objects are assigned to the clusters so the distance from each object to the central vector is minimized. The values of this central vector are called centroids, and they may or may not be part of the objects assigned to their cluster. There are different mechanisms of measuring the distance from a centroid to the objects belonging to its cluster. The most used are Euclidean, Manhattan and Minkowski distance [39].

Euclidean distance is one of the most commonly adopted distance measures. It represents the length of the line segment that connects two objects in a n-dimensional space. It is calculated with the following expression:

$$d_{Euclidean}(p,q) = d_{Euclidean}(q,p) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \tag{2.1}$$

Manhattan Distance is the sum of absolute differences between points across all the dimensions. It is used to calculate the distance between two data points in a grid-like path, and is

expressed in equation 2.2

$$d_{Manhattan}(p,q) = \sum_{i=1}^{n} |p_i - q_i| \tag{2.2}$$

The difference between Euclidean and Manhattan distance is shown in **Figure 2.1**

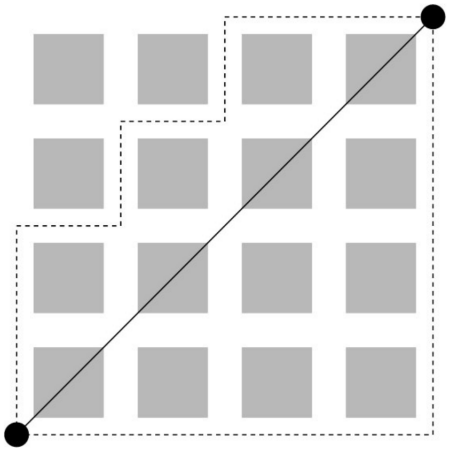

Figure 2.1: Euclidean distance (solid) versus Manhattan distance (dashed) [8]

Finally, the general expression of both Euclidean and Manhattan is Minkowski distance, which is expressed with a general exponent q. When q=1, it represents Manhattan distance, and when q=2 it represents Euclidean distance. The formula is shown in equation 2.3

$$d_{Minkowski} = \sqrt[q]{\sum_{i=1}^{n} |p_i - q_i|^q} \tag{2.3}$$

Once the different measuring mechanisms have been defined, the most common algorithms for centroid-based clustering will be described. These are K-Means, K-Medoids and CLARA.

K-Means

K-Means algorithm is a clustering technique which divides a dataset in k different clusters, each of them having an almost equal number of points. Each cluster is represented by a centroid point, which will be the average of all of the points in the set.
K-means algorithm can be treated as two-phase approach: in the first phase; a k number of centroids are randomly chosen depending on the assigned k value and the cost function is calculated. The cost function is generally expressed as the sum of the Euclidean distance of each point to its centroid. The second phase involves determining the new centroids for which the cost function of the algorithm decreases. The process loops in finding new centroids until the property of convergence is met. [40]

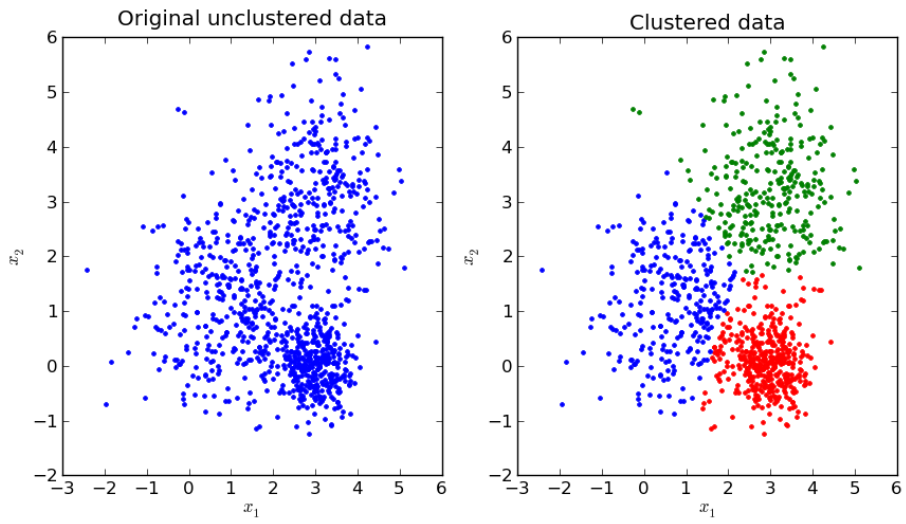

Figure 2.2: Example of k-mean data clustering with k=3 [3]

K-means is a very popular clustering algorithm and has several advantages: it is quick, it can sustain large amounts of data with good results and it reduces intra-cluster variance [40]. However, it is very sensible to the presence of noise and outliers in the data and the fact that the centroids are not actual points of the dataset produces a loss of interpretability in the final results.

K-Medoids

K-medoids is a similar clustering technique to k-means but improves its interpretability. This means that, in this algorithm, the centroids are actual points of the cluster they represent. As well as in k-means, the number k of clusters has to be selected before the algorithm starts. In this case, the cost function is calculated in the form of a dissimilarity matrix using Manhattan distance. The algorithm used in k-medoids clustering is called Partitioning Around Medoids (PAM). It contains the following steps:

1. Random points in the dataset are selected as medoids.
2. Clusters are formed assigning each point of the dataset to its closest medoid.
3. The cost function is calculated for each cluster.
4. Medoids are swapped with a different point from its cluster and the cost function is calculated again. This process is repeated for each point of a same cluster.
5. The process is repeated until convergence is reached.

Figure 2.3 shows the difference in centroid selection between k-means and k-medoids clustering algorithms .

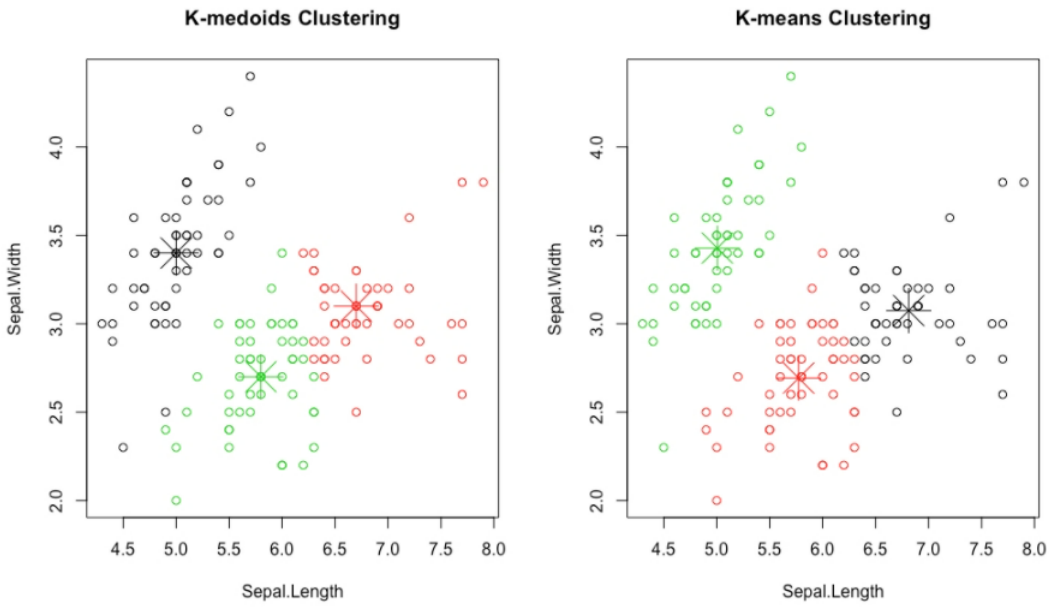

Figure 2.3: The results of k-means clustering vs k-medoids [12]

K-medoids clustering is popular due to its simplicity and ease of implementation. Other advantages are that PAM algorithms are rapid and converge in a fixed number of steps, and, in addition, it is less sensitive to noise and data outliers than other clustering algorithms. However, the initial random assignment of medoids can cause that different runs on a same dataset may obtain different results.

CLARA

Running k-means or k-medoid algorithms on a large dataset can be a computationally expensive task, and this is the reason for the use of Clustering Large Applications (CLARA). It is an extension of k-medoids, but it uses a sampling approach to handle large sets of data. Instead of working with the whole dataset, the algorithm chooses a smaller sample as representative of the data and applies PAM in order to determine the different clusters. This process is repeated for a predefined number of times and returns the best clustering as an output.

The great strength of CLARA is the ability to handle large datasets. However, the efficiency of the clustering depends on the chosen sample size. A small data sample will be computionally accessible, but the quality of the clustering will decrease. Additionally, a good clustering based on samples may not necessarily represent a good clustering of the whole data set if there is presence of bias. **Figure 2.4** shows an example of CLARA applied to a large data set.

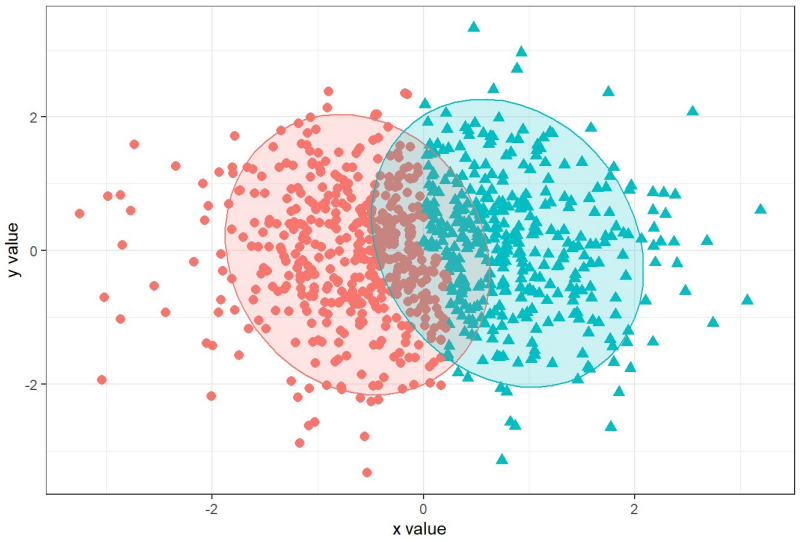

Figure 2.4: Example of CLARA in a large data set [1]

2.2 Time series aggregation in energy optimization

More than just accuracy, solving time and complexity are key elements that needs to be taken into account when designing optimization models [32]. Shorter solving times generally increase the practicality of the model, and some applications even need a certain solving speed. In order to accomplish reduced computational efforts, different approaches are available. Some applications can accept a sub-optimal solution and simply stop the optimization when a certain solving time is reached. A different procedure is to reduce the complexity of the model, by eliminating certain variables or constraints. Finally, it is also possible to reduce the dimensionality of the problem. A way of doing this is lowering the time resolution of the input data and, for example, considering input data every 6 hours instead of in an hourly basis. This last approach is not accurate for energy systems, as a high resolution is needed due to the possible fluctuations of input variables such as demand or temperature in a same time step.

This is the reason why temporal aggregation is a common approach in order to lower computational effort and accomplish reduced solving times. In this context, clustering algorithms appear as an accepted procedure to perform the time series aggregation. According to the revised literature, the two most common temporal aggregation techniques are the use of *typical periods* and *typical time steps* [24].

Typical periods

A way of reducing the number of time steps in a data set is by representing the time series by a set of periods consisting of consecutive time steps. The most common period lengths are days, receiving the name of *typical days* or *representative days*, as many input time series follow a daily pattern, but other period lengths, such as weeks [25] or even months [29], have also been used in certain studies .

Typical time steps

Another way of performing time series aggregation is by selecting a certain number of time steps that will represent the whole set. This technique allows a time series representation using a fewer number of time steps than when using the typical periods approach. However, its use is not recommended for models that include storage systems, as they do not perform well with time-linking constraints [32].

2.2.1 Performance evaluation

In the previous chapter, clustering algorithms have been introduced as a way of performing time series aggregation in order to lower the amount of input data and reduce solving time. In some cases, the model cannot be solved without the use of time series aggregation, so it is not possible to obtain results of the optimization without reducing the amount of time steps. In this cases, the analysis of the influence of the time series aggregation is performed only on the input variables. This is called *a-priori* analysis. When the model is solvable in a reasonable amount of time, and therefor, the influence of the time series aggregation on the optimization outputs can be observed, it recieves the name of *a-posteriori* analysis.

A-priori analysis

Most of the literature consulted on time series aggregation focus their research on a-priori analysis. This allows to obtain results without needing to run an optimization, which may be time consuming, and, in some of them, unfeasible. The use of the k-medoid algorithm in this thesis is based in [16], where this clustering technique is used in reducing a full year of demand data to a few representative days. Dominguez et al. [16] uses ELDC (Error in Load Duration Curve) as a key performance indicator and concludes that a selection of 10 typical days is sufficient for a good representation of the input data in BESs models. Some studies [24, 38], have focused their research in finding out if a priori analysis is sufficient in order to evaluate the performance of different time series aggregation techniques, including clustering algorithms. Both studies conclude that, in general, this is not possible.

A-posteriori analysis

In some cases, if the optimization model is not too complex, it is possible to evaluate the effects of the time series aggregation in the outputs of the actual optimization program. Studies like [27, 31] show that it is possible to reduce up to 90% solving time obtaining relative errors for the objective function of just 2%. Pinel et al.[32] makes a more extended analysis and evaluates the effects of different time series aggregation techniques on the outputs of a Zero Emission Neighbourhood model. The study concludes that clustering algorithms perform more precisely than other techniques in the used model. In the reviewed literature, most key performance indicators used are regression metrics as Root Mean Squared Error (RMSE) [32] or Sum of Squared Errors (SSE) [38]. In this thesis, other metrics, including linear correlation, will also be evaluated.

Along the consulted literature, no evaluations of the effects of time series aggregation applied to approximate MPC applications have been found. Therefore, this work can serve to introduce this topic and analyze the consequences of the use of a clustering algorithm not only on the outputs of an optimization model but also in the results of a rule mining process via a ML algorithm.

3 Use case and approximate MPC structure

In this chapter, the use case and the approximate MPC process used in this thesis will be described.

3.1 Use case

The object of study is a former military hospital, which is being transformed into an office and laboratory center called FUBIC [28]. FUBIC has a HP system in charge of covering the heating and cooling demands of the building, which will be the use case of this thesis.

3.1.1 HP system

The system used to cover FUBIC's heating and cooling demands is pictured in figure 3.1.

This system consists in an HP, an auxiliary electric heater, a heat exchanger and three energy storages: two buffer storages, hot (TES) and cold (CTES), and an ice storage. The whole system is electricity based only. The HP can extract heat from the ambient air, by means of a free cooler, as well as from the ice storage, and even operate reversibly as a chiller. Both buffer storages add flexibility to the system. In order to determine the optimal configuration of the system, 6 different operating modes are defined:

0. The system is off: neither cooling nor heating mode
1. Heating mode: HP extracts heat from outside air via the free cooler
2. Heating mode: HP extracts heat from the ice storage
3. Passive cooling mode: via heat exchanger using the free cooler as heat sink
4. Passive cooling mode: via heat exchanger using the ice storage as heat sink
5. Active cooling mode: via HP using free cooler and ice storage as heat sink

The electric heater can operate independently from the heat pump at any time, charging the TES even if the HP is functioning in a cooling mode. For the storages, two different charging and discharging cycles are considered. On the one hand, the two buffer storages will be constrained with daily cycles. This means that, at the end of the day, the state of charge (SOC) of the storage has to be back to its initial value at the start of that day. The

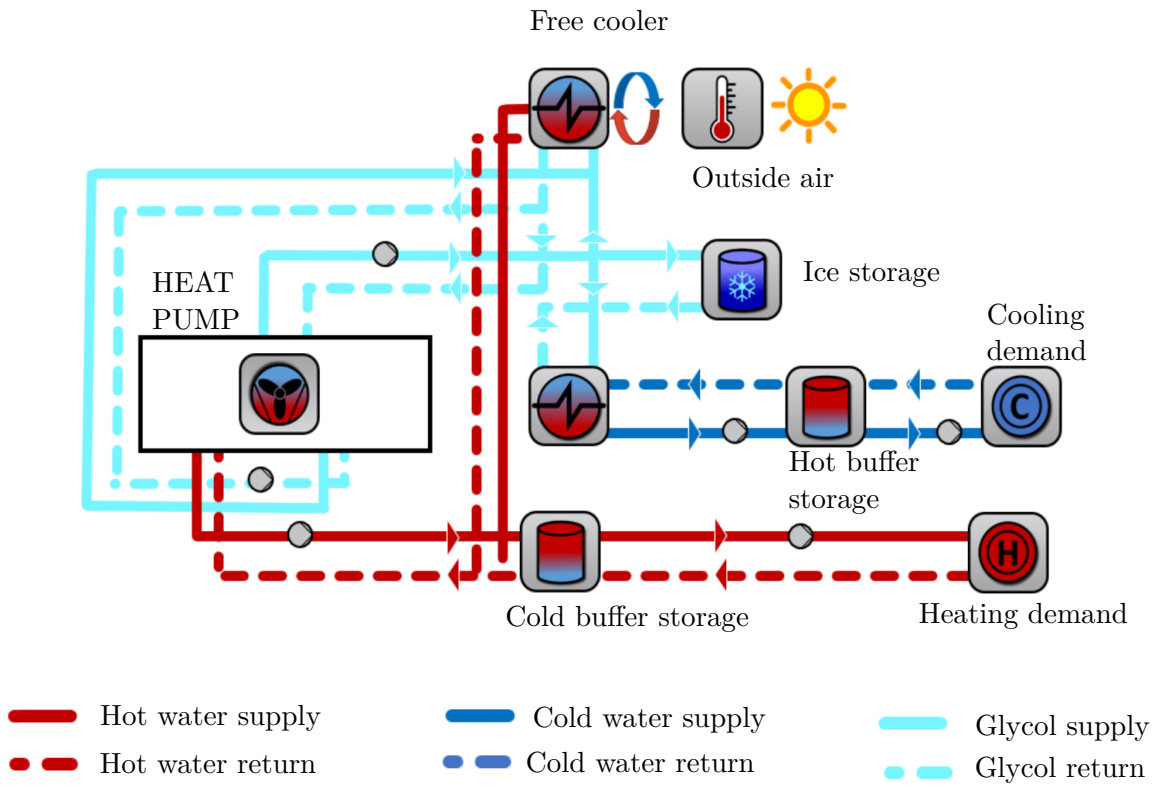

Figure 3.1: HP system of FUBIC

volumes of TES and CTES buffer storages are 40 and 20 m^3 . These small volumes allow daily cycle constraints without limiting the operational possibilities of both storages. On the other hand, the ice storage has a volume of 500 m^3, which allows to explore the possibilities of a yearly cycle constraint. A large charging and discharging cycle enables the integration of longer predictions and increases the number of possible control strategies of the HP system. Additionally, the ice storage provides the possibility of connecting the heating and cooling system. When operating in mode 2, the HP charges the ice storage by extracting heat from it to cover the heating demand. This charging of the ice storage can be used later by the HP when operating with mode 4, discharging the ice tank in order to cover cooling demand.

The Coefficient of Performance (COP) is the ratio of useful heating power to electricity required. It is the most common efficiency metric for HP systems, along with the Energy Efficiency Ratio (EER), which links the provided cooling power with the energy needed to run the system. For this work, both COP and EER of the system will be defined as a quality grade of the Carnot efficiency, which can be expressed only with the temperatures of the heat source and sink. These temperatures will be defined as the arithmetic mean of the supply and return of the heat sink (hot) and heat source (cold). A grade of 50% of the Carnot efficiency

is a common value for HP applications [5].

$$COP = 0.5 \cdot \frac{\frac{T_{supply,hot}+T_{return,hot}}{2}}{\frac{T_{supply,hot}+T_{return,hot}}{2} - \frac{T_{supply,cold}+T_{return,cold}}{2}} \tag{3.1}$$

$$EER = 0.5 \cdot \frac{\frac{T_{supply,cold}+T_{return,cold}}{2}}{\frac{T_{supply,hot}+T_{return,hot}}{2} - \frac{T_{supply,cold}+T_{return,cold}}{2}} \tag{3.2}$$

The return and supply temperatures of heat sources and heat sinks are defined in the following way:

- The supply temperature of the hot water in order to cover the heating demand is always, and the return temperature is 35 °C
- The supply temperature of the cold water in order to cover the heating demand is always 8 °C, and the return temperature is 14 °C.
- A Δ= 8 K is considered for the free cooler. Therefore, when the HP is heating or cooling via the free cooler (mode 1 or 4), there will always be a 6 K difference between the return temperature and the outside air temperature.
- When heating using the ice storage (mode 2), the supply and return temperatures of the glycol are −3 °C and 0 °C, respectively.
- When cooling with the ice storage (mode 3), the supply and return temperatures of the glycol are 6 °C and 12 °C, respectively.
- When the HP is operating as a chiller (mode 5), the free cooler is used as the heat sink. In this case, the supply and return temperatures of the hot circuit are 45 °C and 40 °C respectively. For the cold circuit, the supply and return temperatures of the glycol going from the HP to the heat exchanger are 6 °C and 12 °C, respectively.

3.1.2 Boundary conditions and thermal demand generation

A precise representation of the boundary conditions and energetic demand is a key factor in order to develop an accurate simulation of an energy system. To do so, a reduced order model of FUBIC was developed using the modelling language Modelica. The building was divided into thermal zones of equal usage and boundary conditions and typical user behavior profiles were assigned to each of them, based on investigations by [37]. Additionally, three different weather conditions were used, obtained from the German Meteorological Service [20]: a normal or average year (normal TRY), a relatively cold year (cold TRY) and a relatively warm one (warm TRY). This made possible to observe how the model behaved in response to different scenarios. These boundary conditions are introduced in the open-source Python tool

TEASER [35], where the model is parameterized and then run in the simulation environment Dymola, obtaining the inputs.

The inputs are presented in the form of different text files, where the values for demand and outside temperature appear for each hour over the course of a year. In order to achieve a higher resolution, the input data is lineally interpolated, obtaining a temporal space of 15 minutes between each time step.

3.2 Approximate MPC structure

Modern control strategies, as MPC, can obtain promising results for BESs. However, the high computational requirements, and reduced knowledge among practitioners limit its application. For this matter, an approximate or offline MPC is introduced in order to obtain control rules for our use case with lower computational effort. The approximate MPC process contains a building model, a clustering algorithm, an optimization program and a RF classifier, all provided for the development of this thesis by the Institute of Efficient Buildings and Indoor Climate, from the EON Energy Research Center, a department belonging to the RWTH in Aachen, Germany. Further details of the system can be consulted in [28]. The different steps involving the approximate MPC process are pictured in Figure 3.2 and described below.

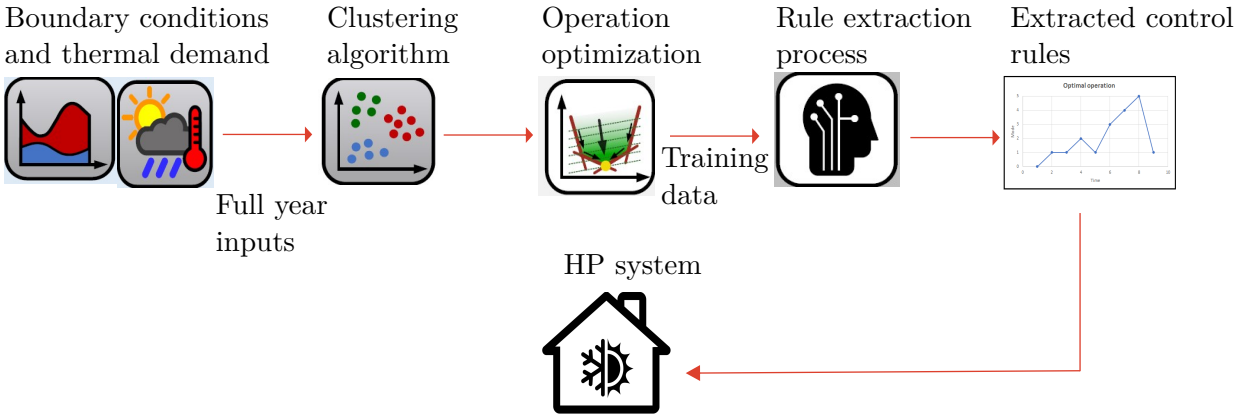

Figure 3.2: Diagram of the approximate MPC process

1. Inputs are generated in the boundary conditions and thermal demand stage.
2. Inputs are clustered by means of a k-medoid algorithm in order to reduce computational effort.
3. The clustered inputs are introduced in the optimization program, where the optimal configuration of the HP is obtained.

4. The optimal configuration of the HP serves as training data for the rule mining process, where control rules are extracted by means of a RF classifier.

5. The extracted control rules are introduced in the HP for it to learn to operate optimally under different scenarios.

3.2.1 Clustering algorithm

In order to generate the typical days that will represent the inputs generated in 3.1.2, a k-medoid clustering algorithm is used. This algorithm receives a data file containing the inputs over the course of a year, and returns a number of typical days. Additionally, it also returns a σ function that allocates each day of the year to its representative cluster. Each medoid has an assigned weight w_d which indicates how often the represented day occurs. The optimization is solved only for the selected representative days and then these results are extended to a full year schedule using σ and w_d. Figure 3.3 shows, as an example, the allocation of each representative day in a full year calendar after a clustering using k=12 medoids.

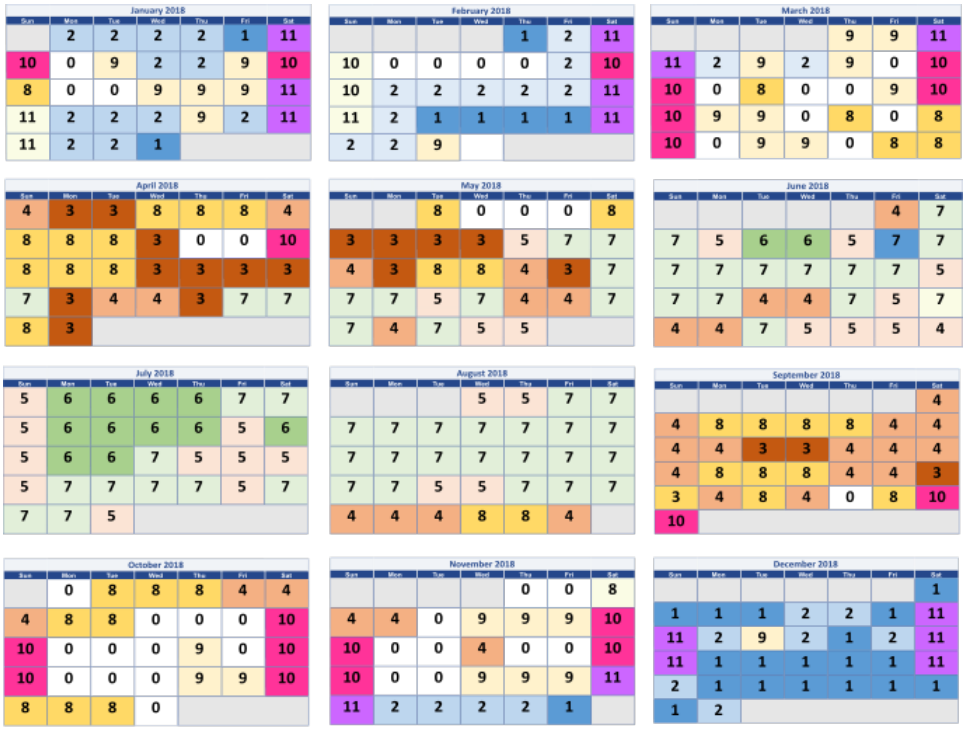

Figure 3.3: Allocation of each representative day in a full year calendar for k=12

It is easy to observe that each day of the year is represented by a typical day, ranging from 0 to 11. This way of representing the output of the clustering also allows to detect some

patterns. For example, representative days 5, 6 or 7 only appear in warmer months, and other typical days like 0, 1 or 2 only show up in the coldest months of the year.

3.2.2 Operation optimization

The optimization program used in this thesis is a MILP developed in Python language, using the package Gurobipy. Gurobi is an online optimization solver which allows to introduce different physical, technical and economical constraints in order to simulate a real life system. The purpose of using this program is obtaining the operational configuration of the HP that will minimize the value of its objective function: the energetic consumption over the course of a year. The program receives the inputs after the data clustering and returns the optimal mode schedule of the HP system for the defined boundary conditions and energetic demand. Additionally, it provides the following information of the state and performance of the system for each time step:

- SOC of the three different energy storages (hot, cold and ice)
- Coefficient of performance (COP) of the system
- EER of the HP when operating as a chiller.
- Charging and discharging powers of the energy storages
- Power of the HP system in its three different modes: air source, ice storage source and chiller
- Power of the auxiliary electric heater
- Optimized value of the OF

The results obtained from the optimization program will be used as inputs for the next step of the approximate MPC process, the rule mining stage.

3.2.3 Rule mining

In the rule mining stage, approximate control rules are extracted from the outputs of the operation optimization. This process is based on an automated data-driven modeling Python tool called AddMo, developed by Ratz et al. [36]. This tool offers a framework for the utilization of data driven machine learning algorithms in black box (BB) modelling and its application in BESs. One of AddMo's objectives is to detect all basic challenges associated with BB modelling, starting from monitoring data and ending with the actual model. The data that is used by AddMo can be divided in two groups: features and signal. Features are the different properties which describe the behaviour of the system. The signal is the variable that the algorithm predicts, i.e, the output of the model. [36]. Input data is also

divided chronologically into two different sets: training and testing periods. In the training period, the algorithm learns the behaviour of the model and the relation between the features and the signal, in order to make a prediction for the testing period. The testing period must contain unseen data, i.e. data that was not included in the training period, to asses the model's capacity to generalize and predict observations. Otherwise, overfitting is guaranteed and it will not be possible to evaluate properly the algorithm's performance.

Process followed by AddMo

The RF classifier makes a prediction of the signal by training it with the different features of the model. The process followed by the algorithm is described below:

Preprocessing: In this step, three different tasks are carried out. The first one is the resolution, were upsampling or downsampling methods are applied in order to obtain the desired resolution of the data set. Next, the data is scaled and normalized, something mandatory in BB models, as the presence of outliers can affect the final output of the algorithm. Lastly, NaNs present in the input data are dealt with through different computations.

Period selection: Period selection shrinks the input data format in the chronological axis by reducing the number of samples [36]. This is done in order to select periods where the data shows the intended behaviour of the model, while other periods are ruled out.

Feature construction: In this step, newly constructed features are added without discarding other features so the BB model has more valuable information, resulting in increased accuracy.

Feature selection: After the feature construction step, the most valuable features are selected. This way, the algorithm rules out the features that do not add helpful information, decreasing unnecessary computational costs.

Sample processing: Sample processing modifies properties of the data set through grouping and shuffling samples [36]. This way, the data is reorganized, without adding or deleting features, but with the purpose of still improving the accuracy of the model.

Model and tool selection: AddMo includes 5 different machine learning algorithms: Support Vector Machines (SVMs), RF, ANNs, Gradient Tree Boosting (GB), and Lasso. AddMo automatically detects which algorithm will provide a better performance for the needed application. Manual selection of the ML algorithm is also possible.

Hyperparameter tuning: BB models contain intrinsic parameters that describe the function or profile of the model. These parameters are called hyperparameters. The objective of hyperparameter tuning is to find the best set of hyperparameters that will improve the performance of the model.

Training, testing and evaluation: Finally, the chosen model is trained with the selected set of hyperparameters on the "tuned data" and is thereafter evaluated [36]. AddMo contains different evaluation measures for both regression and classification purposes.

AddMo was originally developed for carrying out regression tasks. However, for this work, a classifier is needed, so the tool was adapted for this purpose. Both ANN and RF could be used as classification algorithms, but we will focus on the use of RF. This algorithm is chosen because it returns similar results for tabular data than ANN, it requires less preprocessing and the training process is simpler [33]. RF algorithms contain a number of parameters that have an influence on the performance of the model. The parameters included in AddMo for RF classification are the following:

- **Nº of estimators:** These are the number of decision trees the RF algorithm uses. A greater number of trees means a deeper learning process, but also a greater computational effort.
- **Maximum depth:** The maximum depth of the decision trees used in the prediction.
- **Maximum leaf nodes:** The maximum number of leaf nodes, i.e, the final level of the decision tree.
- **Global maximum nº of evaluations:** The maximum number of evaluations the model can perform before obtaining a solution.

4 Method

In this chapter, the method to evaluate the influence of the use of typical days on the results of the approximate MPC will be described. The first step will be to perform an a-priori analysis, i.e, observing the influence of the time series aggregation on the representation of the input curves. Then, the a posteriori analysis will be performed, where the effects of the use of profile days on the results of the operation optimization and the control rules extracted from them will be evaluated.

4.1 A priori analysis

Most of the literature consulted about the influence of time aggregation in optimization models focus only on the effects on the input data. In this section, the focus will be set in this kind of analysis. The three input variables that will be analyzed are the following:

- Outside temperature (ºC)
- Cooling demand (kW)
- Heating demand (kW)

This analysis will be performed to see the differences between how the use of typical days affects the different magnitudes present before and after the operation optimization. It is important to observe the effects on the input variables to compare them with the effects on the outputs of the optimization process. The following key performance indicators (KPIs) are used for this evaluation:

Error in Load Duration Curve

Dominguez et al. [16] introduce Error in Load Duration Curve (ELDC) as a metric for evaluating the accuracy of the clustering algorithm in reproducing input curves. It can be defined as the accumulated relative error between the original data (y_0) and the clustered data (y_c) (see equation **??**.

$$ELDC = \frac{\sum_{i=1}^{n} y_0(t) - y_c(t)}{\sum_{i=1}^{n} y_0(t)} \tag{4.1}$$

Where n is the number of values.

Root mean square error

Root mean square error (RMSE) is a common measure of the differences between observed and predicted values. It has already been used in some studies [31, 32] to compare data before and after performing time series aggregation. One of the benefits of RMSE over other error metrics, as Mean Average Error (MAE), is the fact that errors are squared before they are averaged, giving relatively high weight to large errors [9]. See equation 4.2.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(y_0(t) - y_e(t)\right)^2}{n}} \tag{4.2}$$

RMSE is sensible to the size of the values present in the data set. In order to avoid this, the value of the RMSE will be normalized so it can be compared for the three different variables. A common normalization procedure is to divide it by the mean of the data set [7]. The final value of the normalized RMSE will be calculated as shown in equation 4.3

$$NRMSE = \frac{RMSE}{\overline{y}} \tag{4.3}$$

Where $\overline{y}$ is the mean of the data set.

Coefficient of determination (R^2)

Most of the evaluation metrics for comparing time series before and after aggregation are based on regression metrics. R^2 is defined as the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model [19]. It measures how well the predicted values match the observed values. Considering the original input curves as the observed values, and the clustered input curves as the predicted values, R^2 can be helpful to measure the similarity between the two time series. Correlation coefficients, as Pearson's r, are also used in comparing time series, but they can lead to incorrect interpretations of the results. Figure 4.1 shows an example of how R^2 is a more robust metric than the square of Pearson's r coefficient.

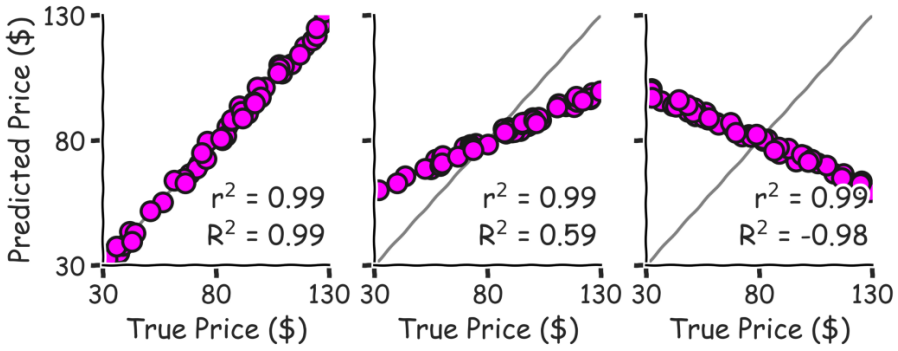

Figure 4.1: Comparison between a correlation coefficient (r^2) and R^2 in a scatter plot [34]

The values of R^2 can range from any negative number to +1. A negative value of R^2 indicates that the predicted time series returns a higher error than the mean of the original time series. The closest the value of R^2 is to 1, the representation of the original time series gets more accurate.

4.2 A posteriori analysis

After evaluating the performance of the clustering algorithm in representing the input curves, the a posteriori analysis is carried out. The aim is to find out in which way the clustering of the inputs has an effect on the operation optimization of the HP system, and on the rules that are extracted from this.

4.2.1 Effects on operation optimization

First, the effects of the use of typical days on the results of the operation optimization described in section 3.2.2 are analyzed. The followed procedure is described below and depicted in figure 4.2.

1. The operation of the HP system is optimized using the full year inputs
2. The operation of the HP is optimized using the clustered inputs described in the previous section. Each simulation is run using inputs obtained by using different number of typical days
3. The results of the simulations with the original data and the clustered data are compared using different KPIs.

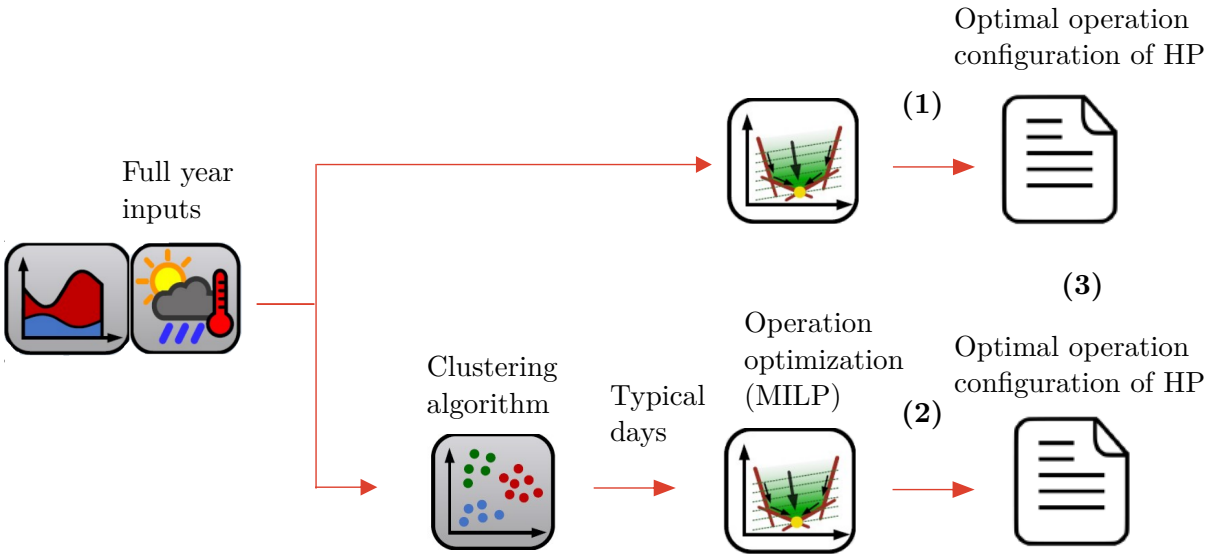

Figure 4.2: Method to evaluate the effects of the use of typical days on the output of the MILP operation optimization

Key performance indicators

The KPIs described in section 5.1 are also used to evaluate the operation optimization when the input data is clustered. In this case, the analyzed outputs will be the following:

- Relative share of usage of each operating mode
- Optimal mode schedule of the HP
- State of charge of energy storages (hot, cold and ice)
- Relative error of the OF value

4.2.2 Effects on rule mining results

The results of the operation optimization are used as training data for the RF classifier. This way, it is possible to visualize the effects of the use of different training data, obtained by optimizing the operation of the HP with different number of representative days, on the control rules extracted by the RF classifier. This analysis will not only focus on the accuracy of the prediction, but it will also evaluate the computational cost of the problem.

The operation optimization returns a great number of output magnitudes. However, many of them may not be useful as features for the RF classifier. The different features are introduced

in the form of a data file containing the value of each feature for every time step for the whole training and testing period. The selected features, for each time step, are the following:

- Optimal mode configuration of the HP for the current time step
- SOC of the three energy storages (hot, cold, ice) at the current time step
- COP of the HP when operating with both the outside air and the ice storage as heat source at the current time step
- EER of the HP when operating reversibly as a chiller for the current time step
- Heating and cooling demand at the current time step
- Outside temperature at the current time step
- Month of the year at the current time step
- Season of the year at the current time step
- Forecast of cooling demand, heating demand and outside temperature for the next two time steps

The steps followed to evaluate the effects of the use of typical days in the rule mining stage are described below and pictured in figure 4.3:

1. The testing period is selected. The optimization program is run with the full year inputs for this period, obtaining the optimal mode schedule for the selected testing period.
2. A training period is selected. The clustering algorithm is run turning the inputs for the training period in a defined number of typical days.
3. The clustered inputs are run through the optimization program, obtaining the optimal HP system configuration.
4. The RF classifier is trained with the outputs of the optimization program obtained in step 3. With this training data, the RF classifier predicts the optimal mode schedule of the selected testing data.
5. The prediction obtained in step 4 is compared to the optimal mode schedule of the testing data obtained in step 1.
6. This process is repeated, using different number of typical days in step 2 for each iteration.

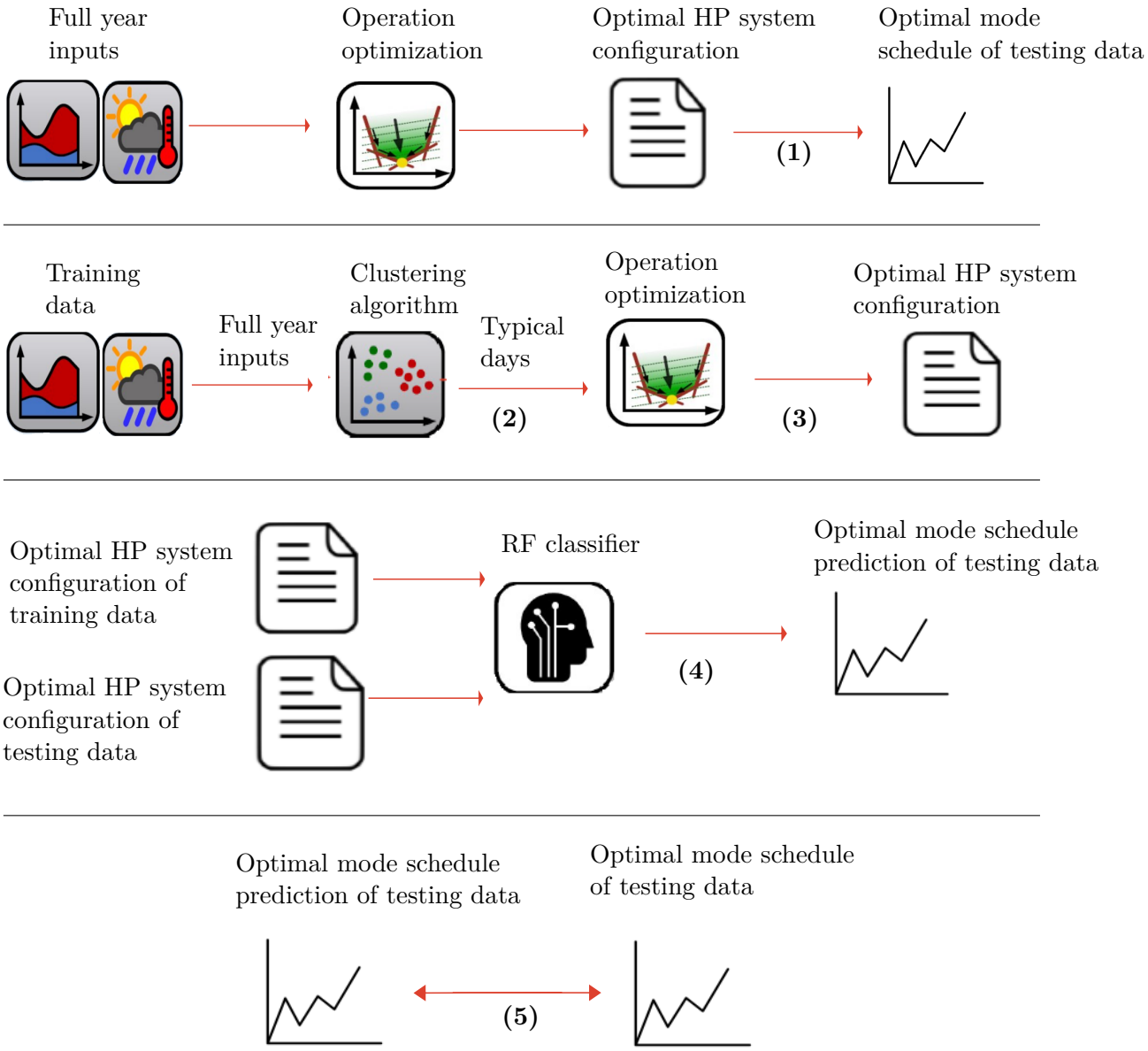

Figure 4.3: Method to evaluate the effects of the use of typical days in the rule mining stage

Key performance indicators

The evaluation of the rule mining process will take into account both the quality of the prediction but also the complexity of the solution. The prediction made by the algorithm is a data column with values ranging from 0 to 5, each corresponding to an operating mode of the HP. This way, it is categorized as a classification problem. Therefore, the quality of the prediction will be assessed using different classification scores.

Confusion matrix

Confusion matrix is a common evaluation metric for Machine Learning classification. It provides a visual way of representing the true and false classifications predicted by the algorithm. Confusion matrices are typically used in binary problems (2 classes), where there are only four possible evaluations for the prediction: true positive, true negative, false positive and false negative. Table 4.1 shows an example of a confusion matrix applied to a binary classification problem.

		True class	
		Positive (P)	**Negative (N)**
Predicted class	**Positive (P)**	True positive (TP)	False positive (FP)
	Negative (N)	False negative (FN)	True negative (TN)

Table 4.1: Example of two-class confusion matrix

This metric can also be applied to a multi-class application, like the problem described for this thesis (see table 4.2) This way, the values of TP, TN, FP and FN are applied for each of the classes present in the problem.

		True class					
		0	**1**	**2**	**3**	**4**	**5**
Predicted class	**0**						
	1						
	2						
	3						
	4						
	5						

Table 4.2: Example of 6 class confusion matrix

Accuracy, precision and recall

The elements of the confusion matrix can be used to calculate different classification metrics. The most common ones are accuracy, precision and recall [10].

Accuracy measures the relation between the number of correct predictions and the total number of predictions.

$$Accuracy = \frac{\text{TP}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4.4}$$

In the case of a multi class problem, precision and recall must be referred for each individual class. **Precision** measures, of the total number of predictions for one class, how many of them are correct.

$$Precision = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4.5}$$

Recall measures how many samples of one class where detected by the classifier .

$$Recall = \frac{\text{TP}}{\text{TP + FN}} \tag{4.6}$$

Choosing a correct metric for a multi class classification problem is a difficult task, especially if there is an unbalanced distribution of the different classes. This problem has been assessed in different studies as in [21, 22, 30]. When dealing with the unbalanced multi class problem, using conventional classification metrics can generate misleading results. There can be problems where at least one of the classes constitutes only a very small minority of the data set. In this cases, a prediction with no observations of these classes may return high accuracy scores while being unable to detect some of the classes which constitute the problem. To avoid this, Mosley et. al [30] suggest the use of Class Balanced Accuracy (CBA), which can be defined with the following formula:

$$CBA = \frac{\sum_{i=1}^{n} \frac{TP_i}{max(TP_i+FP_i,TP_i+FN_i)}}{i} = \frac{\sum_{i=0}^{n} min(Precision_i, Recall_i)}{n} \tag{4.7}$$

Here n is the number of classes. A highly unbalanced problem can lead to very low values of precision and recall for the least frequent classes, reducing notoriously the value of CBA. To avoid this, a complementary metric will also be used, where each class will be weighted according to its frequency of appearance in the testing data. This metric will be named Weighted Class Balanced Accuracy (WCBA), the formula for which is defined below:

$$WCBA = \sum_{i=1}^{n} min(Precision_i, Recall_i) \cdot w_i \tag{4.8}$$

Where

$$w_i = \frac{\text{Samples of class n}}{\text{Total number of samples}} \quad , \quad \sum_{i=1}^{n} w_i = 1 \tag{4.9}$$

The KPIs previously described in this section refer exclusively to the performance of the algorithm and the accuracy of its prediction. However, there are other aspects to consider. One of them is how complex and time consuming can it be to obtain an accurate prediction. The other is how much data is needed in order to perform an adequate training of the model. In order to evaluate this, two additional KPIs are introduced.

Computation time and training period

The parameters described in section 3.2.3 have an effect on the performance of the RF classifier but also on the complexity of the problem. When using large values for these parameters, the algorithm makes a higher number of evaluations and chooses between a

higher number of possible solutions. However, this results in higher computational effort and solving times . The time needed to train the model and choose an specific solution will be called 'Computation time', which will be shown for each experiment among the classification metrics described in this section.

Another way of evaluating the complexity of the problem is figuring how much training data is needed to obtain an accurate prediction. The larger the set of training data is, it requires the study of more weather conditions and more user demand profiles, which increases the overall complexity of the problem. For this reason, the last KPI will be the amount of data used in the training of the model, which is defined as the length of the training period, i.e, the number of samples for each feature that make up the training data.

5 Results

In this chapter, the results obtained by following the method described in chapter 4 will be displayed.

5.1 A priori analysis

In the a priori analysis, the inputs used for the operation optimization are compared before and after the clustering. The analysis is performed based on the three weather conditions: cold, warm and normal TRY.

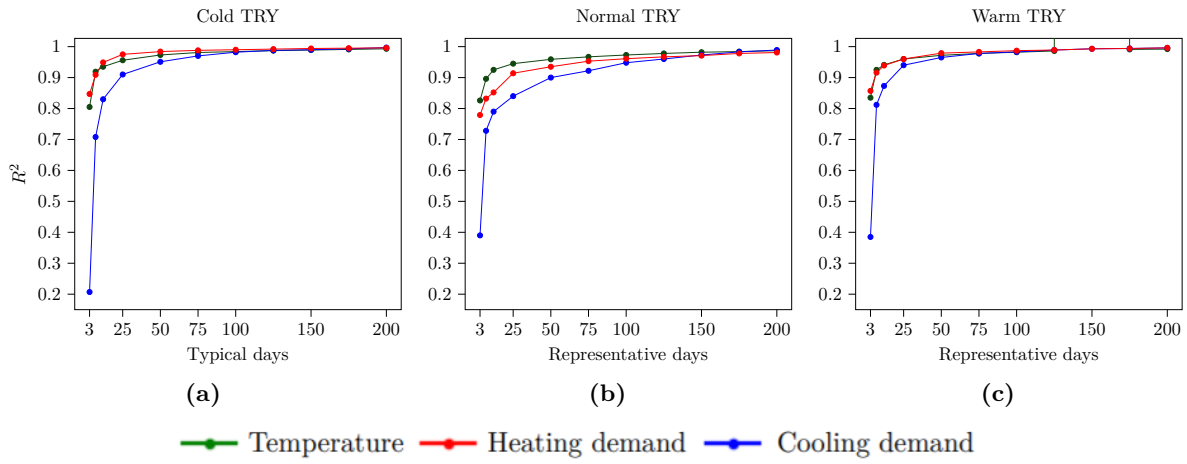

Figure 5.1: R^2 in - obtained by comparing the original and clustered input curves for cold TRY (a), warm TRY (b) and normal TRY (c) for different number of representative days

Figure 5.1 shows the values of R^2 when comparing the original and the clustered input curves for the three weather conditions. In all cases, the results reflect a clear trend: as the number of selected typical days increases, the correlation gets stronger. For 3 representative days, the values of R^2 for cooling demand is noticeably worse than in the case of the other two input variables, but it grows rapidly for 12 typical days and higher. For higher number of representative days, the correlation between the original and the clustered input curves is almost perfect, reaching values of R^2 higher than 0.9.

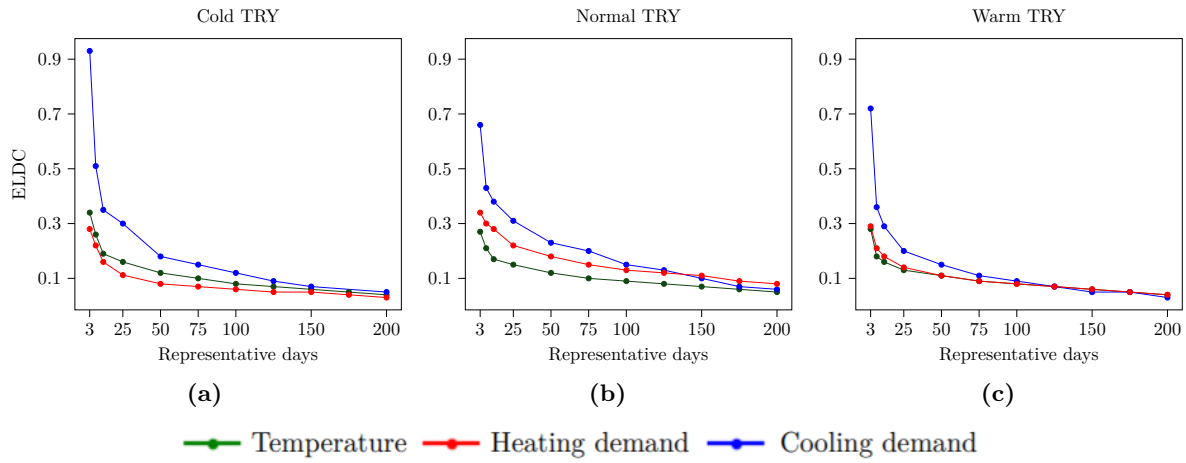

Figure 5.2: ELDC in - obtained by comparing the original and clustered input curves for cold TRY (a), warm TRY (b) and normal TRY (c) for different number of representative days

As well as in figure 5.1, the ELDC index shows that the more days are used, the more accurate becomes the generated input curve with regard to the original one. Both temperature and heating demand, for all weather conditions, return values of ELDC just above 0.3 for only 3 typical days, and progressively decrease as the number of days is raised. However, as was the case for figure 5.1, cooling demand shows the larger error metrics of the three input variables, especially for 12 typical days or lower, returning values up to 0.9. As the number of days is increased, the ELDC index for cooling demand becomes more similar to those of the other variables.

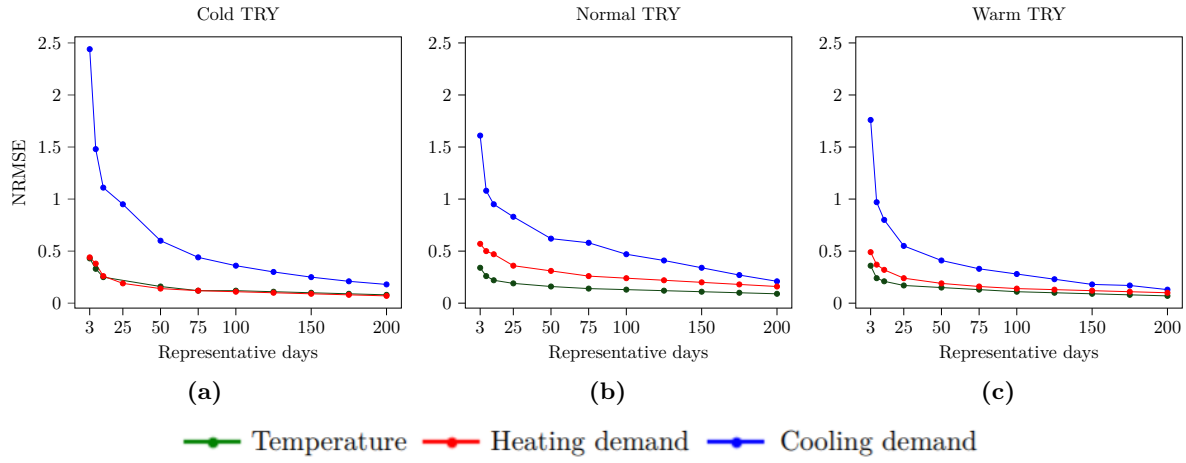

Figure 5.3: NRMSE obtained by comparing the original and clustered input curves for cold TRY (a), warm TRY (b) and normal TRY (c) for different number of representative days

The RMSE index shows that the performance of the clustering algorithm is clearly poorer for the cooling demand than for the other two input curves. In the case of temperature and heating demand, the values of NRMSE are lower than 0.5 for any number of typical days. This does not happen for cooling demand, where it takes a selection of 100 typical days to reach the same performance than for the other two input variables.

The three selected KPIs show similar results for the performance of the clustering algorithm. As expected, a higher selection of typical days results in a greater similarity between the original and clustered input curves. In general, cooling demand is the input variable which returns higher error metrics. This is especially remarkable for the NRMSE index, where the results for cooling demand are significantly worse than in the case of the other two input curves. It is also possible to observe this trend for R^2, and in a lesser degree, for ELDC, especially when using very low numbers of selected typical days.

5.2 A posteriori analysis

The a posteriori analysis evaluates how the use of typical days has an effect on the outputs of the operation optimization and on the rules extracted from them in the rule mining process.

5.2.1 Effects of typical days on operation optimization of HP system

The first step of the analysis is to evaluate how the use of typical days has an effect on the optimal mode configuration of the HP system. After an operation optimization, it is possible to observe the relative usage of each of the operating modes of the HP, and detect if this relative usage changes when using different number of typical days as inputs for the optimization. This is pictured in figure 5.4.

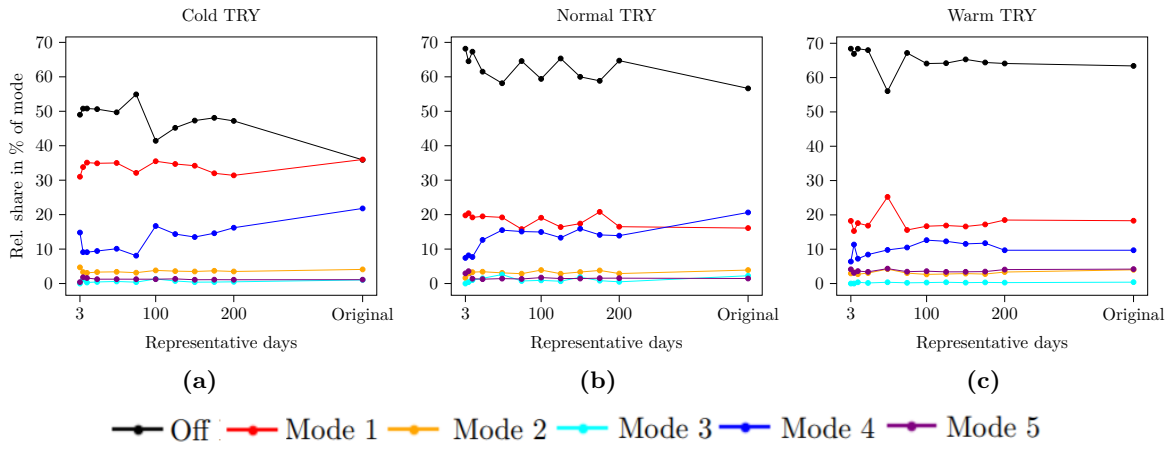

Figure 5.4: Relative share in % of the different modes of the HP in an operation optimization using different number of typical days for cold TRY (a), normal TRY (b) and warm TRY(c)

The first detail that stands out when looking at figure 5.4 is the great imbalance in the use of the different modes. For most of the time, the HP is switched off, with mode 1 and mode 4 being the next most frequently used modes. In all cases, mode 2, 3 and 5 are the least utilised ones, with relative shares lower than 5 %. The optimal mode configuration for the cold TRY shows a smaller imbalance than for the other two weather conditions. The relative share of mode 0 (HP is off) is, except for 75 typical days, lower than 50 %, and the relative share of mode 1 is higher than 30 % in every case. Mode 4 has also a greater usage than in the case of the other two weather conditions. Another appreciable aspect is the fact that the relative share of the different operating modes tend to get more balanced as the number of typical days is increased. This trend occurs especially in the cold TRY, being less noticeable in the normal TRY and almost imperceptible in the warm TRY. In the normal TRY, mode 4 gains more importance in the relative share as the number of typical days used in the optimization increases, while the use of the other operating modes remain almost constant. The HP's optimal configuration pictured in figure 5.4 is used as a training feature in the rule mining process, and it remains to be seen how the changes in the mode distribution caused by the use of typical days affects the control rules extracted from the optimization results.

The next part of the analysis consists of evaluating the effects of the use of typical days in the charging and discharging cycles of the three energy storages. These cycles provide a decent description of the operation of the system, as the SOC of the energy storages, for each time step, is conditioned by the energetic demands as well as the operating mode of the HP. This analysis is performed for the three weather scenarios.

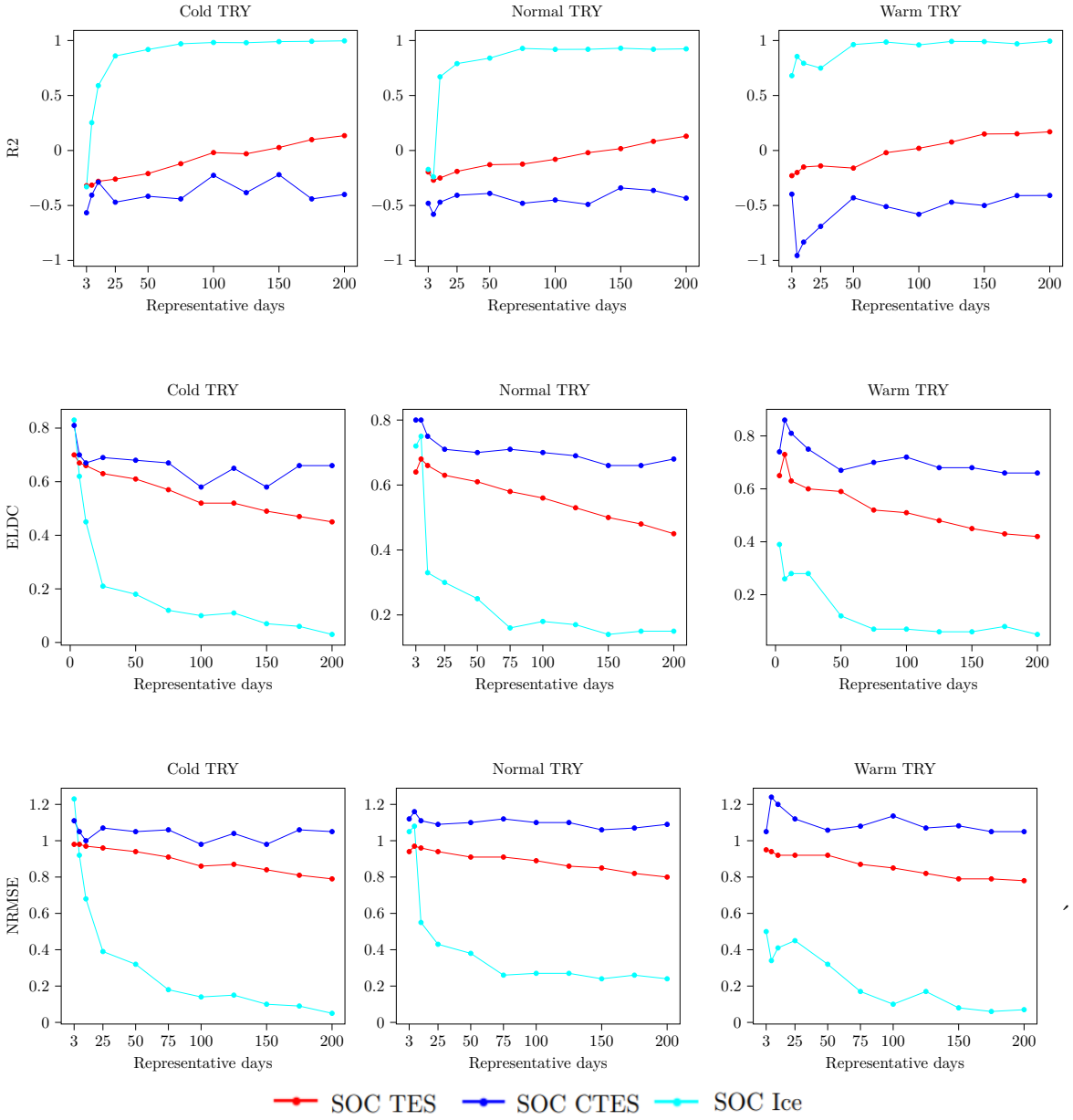

Figure 5.5: R^2, ELDC and NRMSE for operation optimization outputs when using different number of typical days

The results for the optimization outputs show some similar trends as the ones obtained in the previous section. It is possible to observe that all of the error metrics, for the three weather conditions, return better results as the number of typical days is increased. Another common aspect among the three displayed KPIs is the fact that the charging and discharging cycle of the ice storage returns lower errors than the ones of the buffer storages. For the cold and normal TRY, the ice storage cycle returns low values of R^2, but they increase rapidly when using 12 typical days or higher. For a higher number of days, the correlation between the

original optimal ice storage cycle and the one generated with the use of typical days is very strong, returning values of R^2 higher than 0.9. For the charging and discharging cycles of the buffer storages, the obtained values of R^2 are lower than 0 for almost any number of typical days, which indicates no correlation between the generated optimal cycles with the use of typical days and the original ones.

Both ELDC and NRMSE show similar results for the three weather conditions. For a number of typical days lower than 12, the three storage cycles show error metrics larger than 1. However, when the number of days are increased, the ELDC and NRMSE values for the ice storage cycle decrease very rapidly, reaching values lower than 0.4 for both indexes. In the case of the cycles of the two buffer storages, the values of ELDC and NRMSE also decrease when raising the number of representative days, but at a much slower rate. Even so, the cycle of the TES returns better error metrics than the CTES for any selected number of typical days. For 200 days, the ELDC index returns values lower than 0.2 for the cycle of the ice storage, while the values for the TES and CTES are approximately 0.8 and 1, respectively.

In order to visualize the charging and discharging cycles of the different energy storages, figures 5.6, 5.7 and 5.8 show the evolution of the SoC of the the three storages for 5 days, in a summer week and in a winter week, obtained when optimizing the system using original inputs and 50 typical days. The dashed lines present in the figures 5.6 and 5.7 indicate the places where the daily cycle constraint must be met.

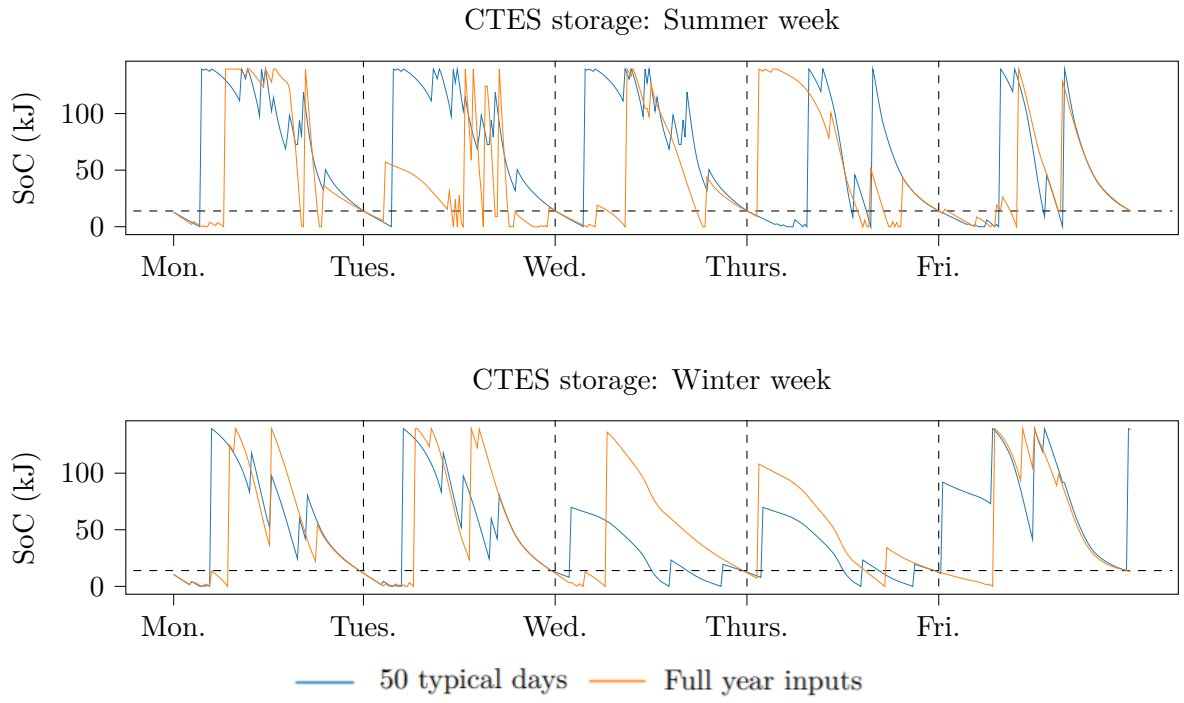

Figure 5.6: Charging and discharging cycle of CTES storage

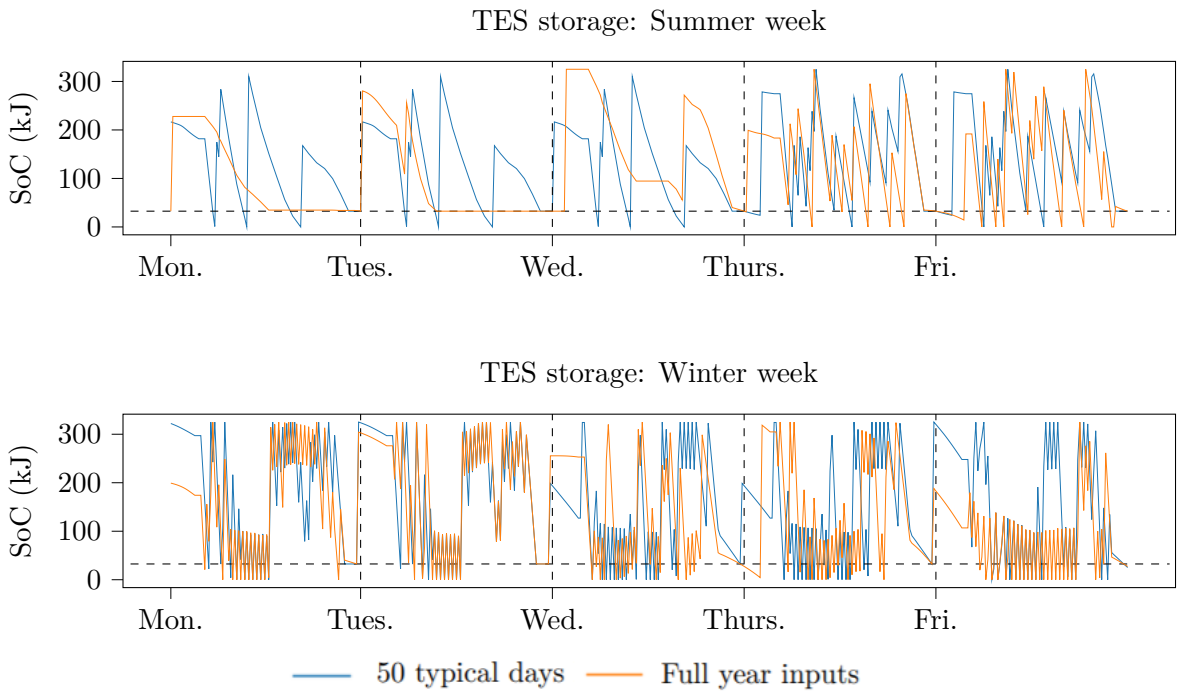

Figure 5.7: Charging and discharging cycle of TES storage

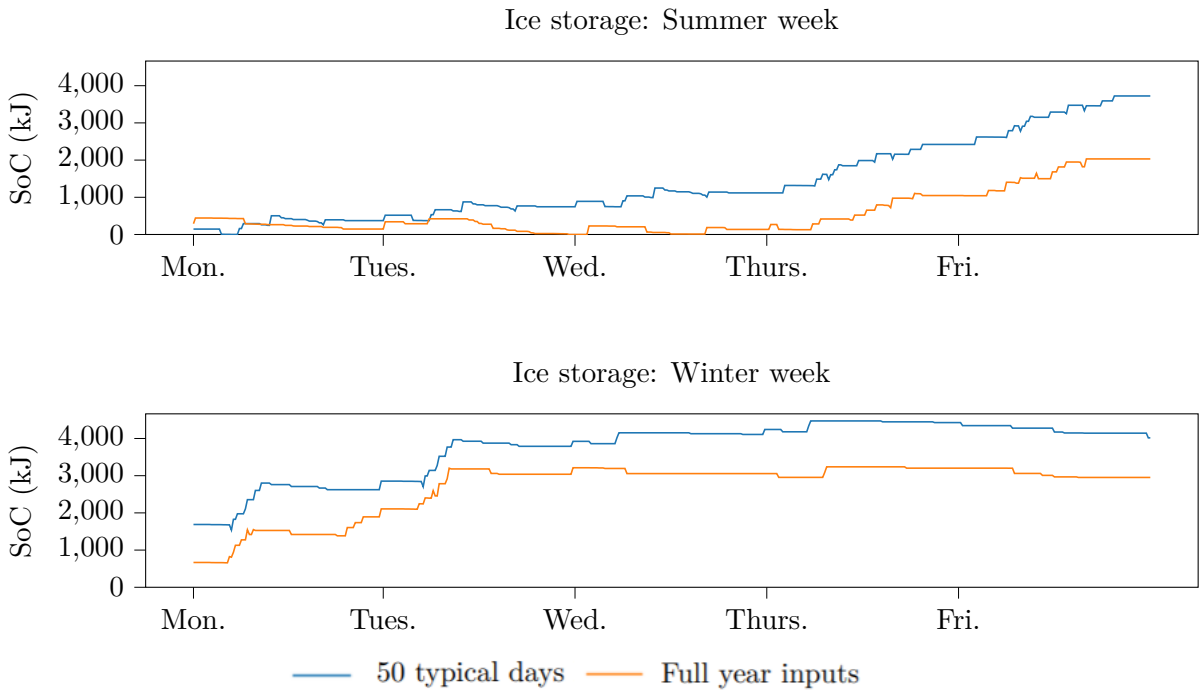

Figure 5.8: Charging and discharging cycle of ice storage

Figures 5.6 and 5.7 show a great variability in the state of charge of both TES and CTES storages, which is difficult imitate when using the typical day approach. This is due to their small volume (40 m^3 for the TES and 20 m^3 for the CTES), causing that small charges or discharges of energy have a great impact over the total SoC. This is the reason why the R^2, ELDC and NRMSE scores, for these two storages, reflect such poor values. The opposite happens in the case of the ice storage. Although there is a greater difference between the SoC for the original input optimization and the one obtained using typical days, both curves follow similar trends. The greater volume of the ice storage (500 m^3) enables a greater capability of absorption of small energy variations, causing a more stable trend in the SoC, which is easier to represent when using the typical days approach.

The last step of this analysis is to evaluate how the use of typical days has an effect on the value of the OF of the operation optimization. In order to do so, the value of the OF was obtained for a simulation with the original inputs, and compared with the ones obtained when using different number of typical days. This was performed for the three weather conditions (see figure 5.9).

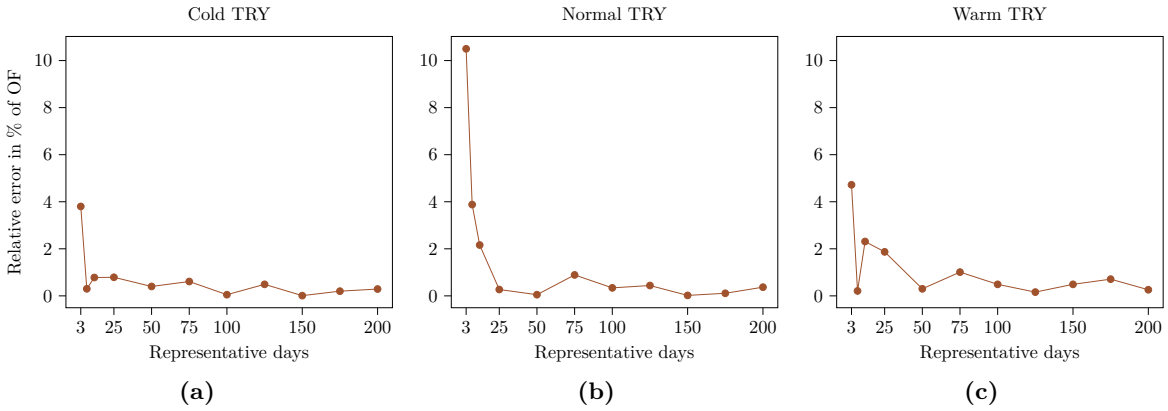

Figure 5.9: ELDC in - obtained by comparing the original and clustered input curves for cold TRY (a), warm TRY (b) and normal TRY (c) for different number of representative days

Figure 5.9 shows an important similar results for the three weather conditions. For a very low number of typical days, the generated value of the OF has a considerable error compared to the one obtained with the original inputs. This is especially remarkable in the case of the normal TRY, where the relative error of the OF exceeds 10 %. However, this error decreases rapidly when increasing the number of typical days. For 25 representative days, the relative error falls bellow 1%, and then fluctuates between an interval of ± 0.5 % when increasing the number of typical days up to 200. This also happens for the cold TRY. For the warm TRY, it takes 50 typical days to obtain a relative error lower than 1 %.

5.2.2 Effects of typical days on rule mining

The effects of the use of typical days on the results of the operation optimization of the HP system have been assessed in the previous section. As these results are used as training data for the rule mining stage, the next step is to evaluate how these effects on the training data affect at a later stage the extracted control rules. As mentioned in section 4.3, the aim of this thesis is to find out if the use of typical days is a valid approach to obtain simplified control rules in an approximate MPC process. In order to evaluate this, both the accuracy of the prediction made by the RF classifier and the time needed to obtain the solution will be taken into account.

The first step is to set a benchmark, which will be used as a reference to compare the different results obtained in the rule mining stage. In order to do so, some conditions will be selected to obtain the best possible performance of the RF classifier. This conditions are the following:

- Training data obtained after an operation optimization with the original inputs, i.e. not using typical days.
- High values for the different parameters of the RF classifier. This will result in a more complex training process, but also in higher computational effort.

As mentioned in section 4.3, the amount of training data is one of the KPIs used to evaluate the complexity of the problem. For this purpose, two different training periods will be used: 18 months and 30 months.

Training period: 18 months

The first experiments are carried out using a training period of 18 months, which, according to [28], is sufficient to capture the behaviour of the optimal configuration of the HP system. However, Maier et al. use a daily cycle condition for the charging and discharging cycle of the ice storage, and highlight the fact that maybe a larger training period is needed when using a yearly cycle. The 18 months training period consists of 12 months of cold TRY and the first 6 months of the warm TRY, while the testing period will be the last 6 months of the warm TRY. In order to obtain a benchmark, the experiments described in table 5.1 are carried out. For each different experiment, the values of the RF parameters are increased in order to see how it has an effect on the accuracy of the prediction, but also in the solving time needed to obtain a solution.

Exp.	Nº of estimators	Max. depth	Max. nº of leaf nodes	Max. nº of eval.
1	100	1000	5000	100
2	200	2000	5000	100
3	300	3000	5000	100
4	400	4000	5000	100
5	500	5000	5000	100

Table 5.1: Values of the RF parameters used for the benchmark predictions for 18 months of training

The results of the first predictions are pictured in figure 5.10. In addition, the values of accuracy, CBA and WCBA are shown for each prediction, along with the solving time needed to obtain the respective solutions.

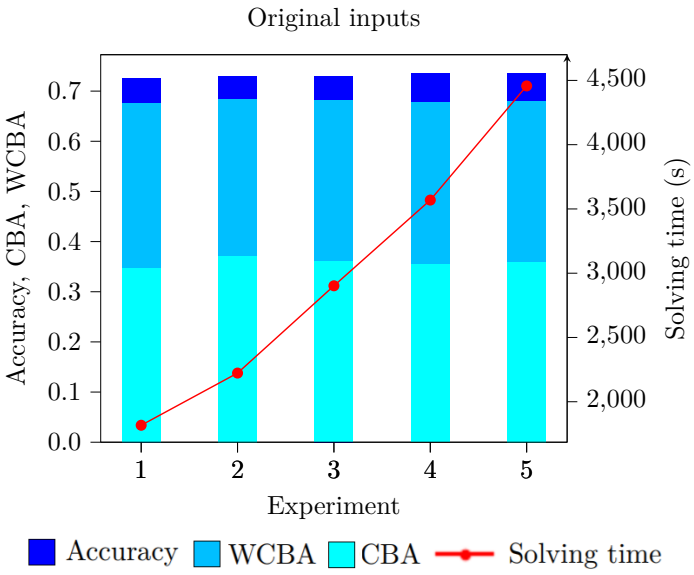

Figure 5.10: Accuracy, WCBA, CBA and computation time obtained for the experiments described in table 5.1 for training data generated with original inputs and training period of 18 months

The first thing that stands out when looking at the previous figure is that the accuracy, CBA and WCBA scores is almost the same for the 5 experiments. However, the time needed to reach the different solutions increases rapidly as the values of the RF parameters are raised. This probably means that the limit of performance of the classifier is reached, i.e. increasing the complexity of the problem does not provide better results for the prediction.

Another significant condition illustrated by figure 5.10 is the great difference between the values of CBA and WCBA for each experiment. This indicates that the algorithm performs very well in predicting the classes with a higher frequency of appearance, but not so well with

the less frequent classes. This is pictured in figure 5.11.

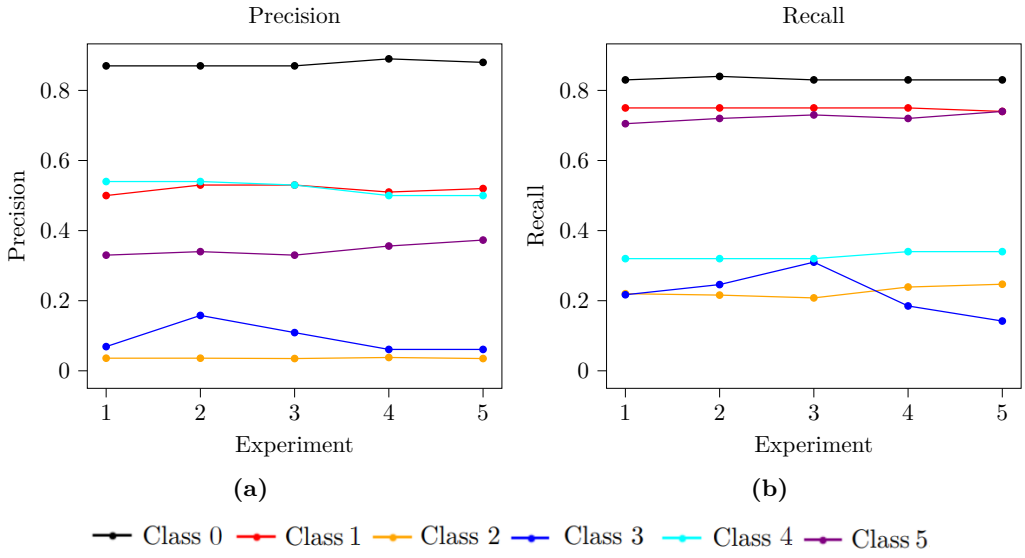

Figure 5.11: Values of precision (a) and recall (b) of each class present in the predictions pictured in figure 5.10

	Cold TRY	Normal TRY	Warm TRY
Off	35.9%	56.6%	63.4%
Mode 1	36%	16.1%	18.3%
Mode 2	4.1%	3.9%	4%
Mode 3	1%	2.3%	0.4%
Mode 4	21.8%	20.6%	9.7%
Mode 5	1.1%	1.5%	4.2%

Table 5.2: Relative share of each mode of the HP for an operation optimization with full year inputs and yearly cycle condition for the ice storage

Figure 5.11 shows that the precision and recall scores for the most frequent classes are clearly higher than for the less frequent ones. This is very noticeable, for the opposite reasons, for class 0 and class 2. Class 0 is by far the one with most samples, for the three weather conditions, and displays values for precision and recall always higher than 0.8. However, for class 2, with a relative share in the training data of lower than 5 %, both precision and recall score are very low. This contrast in the accuracy of the algorithm in predicting the most frequent and less frequent classes causes the big gap between the values of CBA and WCBA. Low values of precision and recall for, for instance, class 2, have a great impact in the final score of CBA, but not so much in WCBA, as the weight corresponding to its frequency of

appearance is very low.

Once the benchmark for the performance of the RF classifier has been established, the next step is to carry out experiments using results from the operation optimization, when using typical days as training data. The analysis carried out in sections 5.1 and 5.2.1 suggests that the use of typical days provides better results when using 12 days or higher. For this reason, the experiments will start using training data obtained from an operation optimization using 12 typical days. For each set of training data, several experiments are carried out, using different values for the RF parameters.

Among the performed predictions, there is some variability in the results. For this reason, only the results which are representative of the aim of the thesis are displayed. Figure 5.12 shows 5 examples of these experiments.

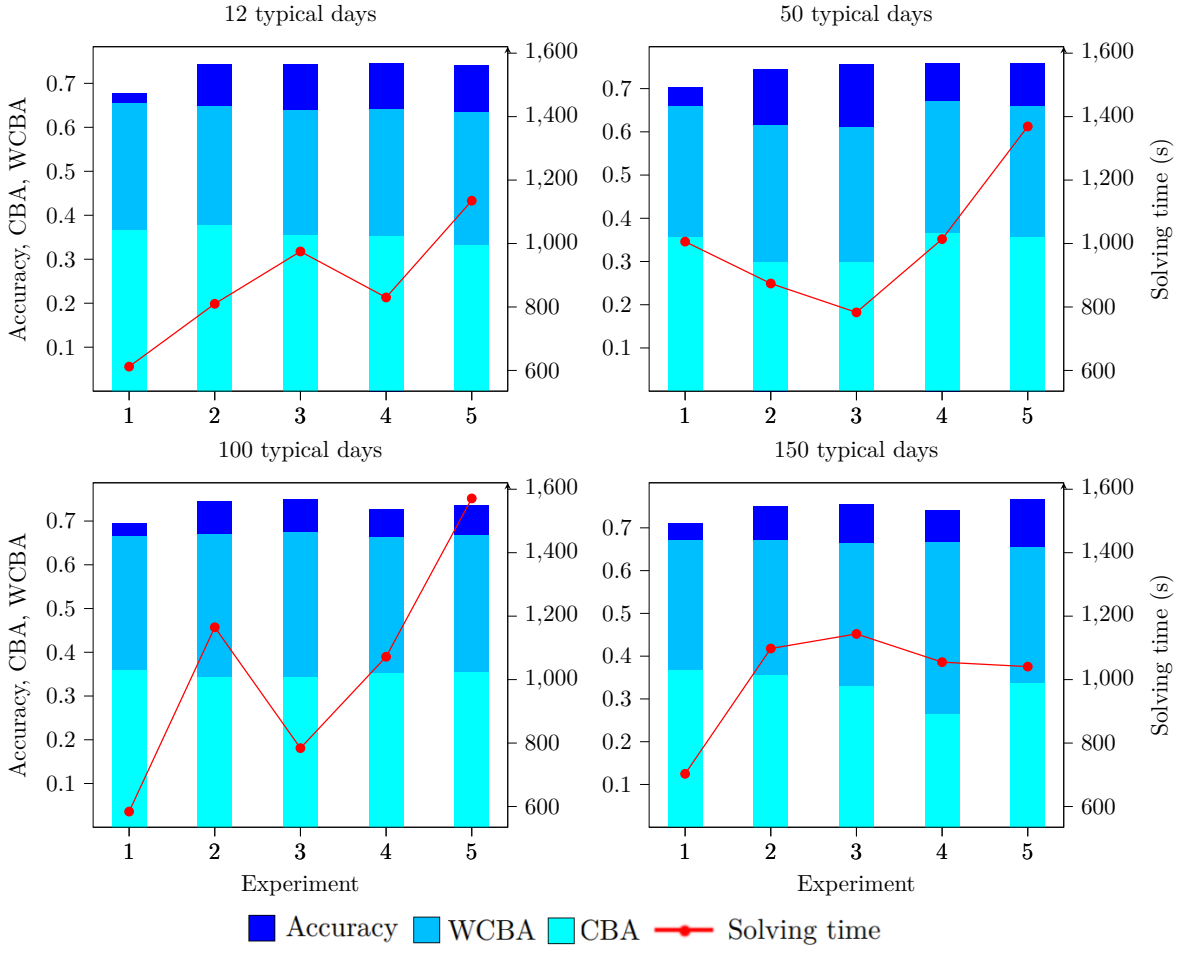

Figure 5.12: Accuracy, WCBA, CBA and solving time obtained for representative experiments using training data generated with different number of typical days for training data of 18 months

Figure 5.12 shows very similar results in terms of accuracy scores as the ones displayed in figure 5.10. However, the times needed to obtain each solution are significantly lower. Using typical days in the operation optimization results in a simplified version of the optimal configuration, but this does not seem to affect in the performance of the RF classifier. In fact, there is almost no difference in the results obtained when using different number of typical days in the training data. As well as in the predictions carried out to obtain the performance benchmark, there is an important difference between the values of CBA and WCBA scores for each experiment, indicating a high accuracy in the prediction of the more frequent classes and a poorer performance with the less frequent classes. However, the results are not very good. The great imbalance in the relative share of the different operating modes causes that predicting accurately just mode 0 and 1 can be enough to obtain decent accuracy scores. This is not reflected by the CBA metric, which fluctuates between 0.3 and 0.35 in the best cases.

A possible explanation for this may be that 18 months of training data is not sufficient to capture the seasonal patterns of the ice storage. Although the charging and discharging cycle of the ice storage was, among the analysed optimization outputs, the one which presented lower error metrics in section 5.2.1, maybe a greater training data set is needed to extract accurate control rules. In order to test this, the same procedure followed for 18 months of training will be carried out for a training period of 30 months.

Training period: 30 months

The training period of 30 months consists of 12 months of cold TRY, 12 months of normal TRY and the first 6 months of the warm TRY. As well as in the previous experiments, the testing period will be the last 6 months of the warm TRY. The first experiments are carried out with the same conditions as the ones described in table 5.1, in order to obtain a performance benchmark for the RF classifier.

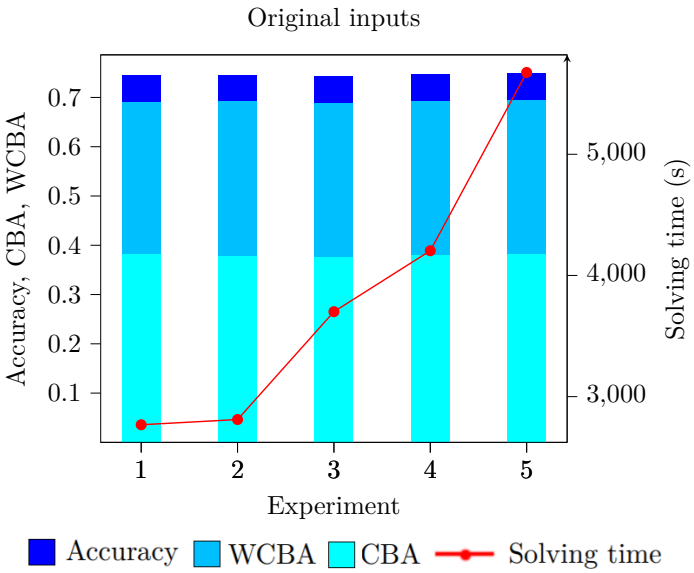

Figure 5.13: Accuracy, WCBA, CBA and solving time obtained for the experiments described in table 5.1 for training data obtained with original inputs and a training period of 30 months

The obtained results are slightly better than for the training period of 18 months. The scores for CBA and WCBA for the 5 experiments almost reach 0.4 and 0.7, respectively, which is an improvement between 5 and 10 % over the previous case. However, the needed solving times are significantly higher, as a longer training period requires more time in the preprocessing and selection of periods and features. As well as for the training period of 18 months, there is considerable gap between the values of CBA and WCBA. Figure 5.14 shows the precision and recall scores for the different classes present in the predictions carried out in the benchmark.

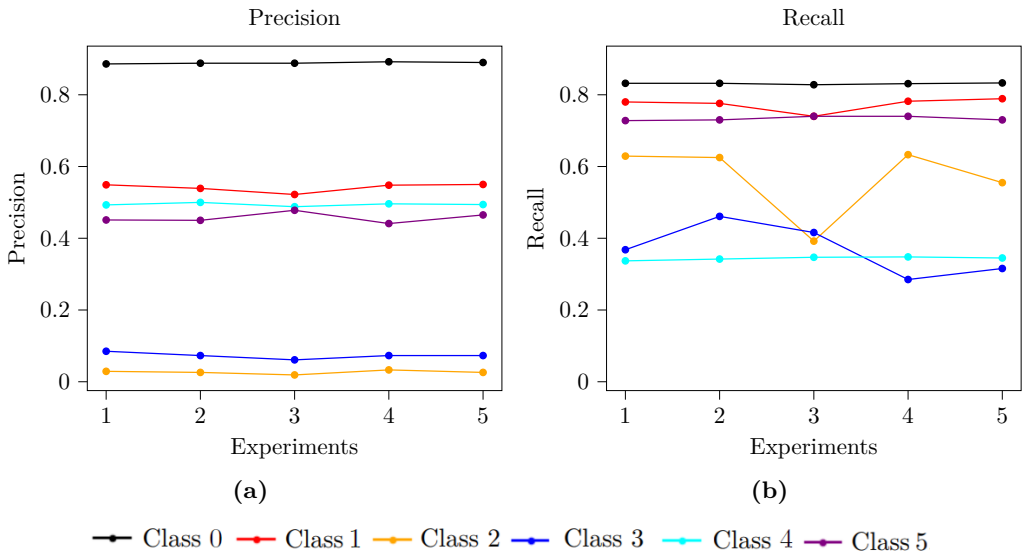

Figure 5.14: Values of precision (a) and recall (b) of each class in the experiments carried pictured in figure 5.13

	Cold TRY	**Normal TRY**	**Warm TRY**
Off	35.9%	56.6%	63.4%
Mode 1	36%	16.1%	18.3%
Mode 2	4.1%	3.9%	4%
Mode 3	1%	2.3%	0.4%
Mode 4	21.8%	20.6%	9.7%
Mode 5	1.1%	1.5%	4.2%

Table 5.3: Relative share of each mode of the HP for an operation optimization with full year inputs and yearly cycle condition for the ice storage

Figure 5.14 shows similar trends but also some improvements over the case for 18 months of training data. There is a significant difference between the precision and recall scores of the different classes according to their frequency of appearance, being class 0, evidently, the one with better results, with precision and recall scores higher than 0.8 for all experiments. However, class 5, although having one of the lowest relative shares (lower than 5 % for the three weather conditions), returns similar precision and recall scores than other classes with higher frequency of appearance, as class 1 and 4. The recall of class 2 and 3, the other least frequent classes, also has undergone an improvement compared to the case of 18 months of training data, increasing from values loer than 0.2 to almost 0.4. This, however, has no effect on the CBA score, as this metric picks, for every class, the minimum value between precision and recall, and the precision scores for these classes are still comparatively low.

Once the benchmark for the performance of the RF classifier for a training period of 30 months has been set, several experiments are carried out, with different values for the RF parameters and using training data obtained by optimizing the operation of the system with different numbers of typical days. The results of the predictions are pictured in figure 5.13.

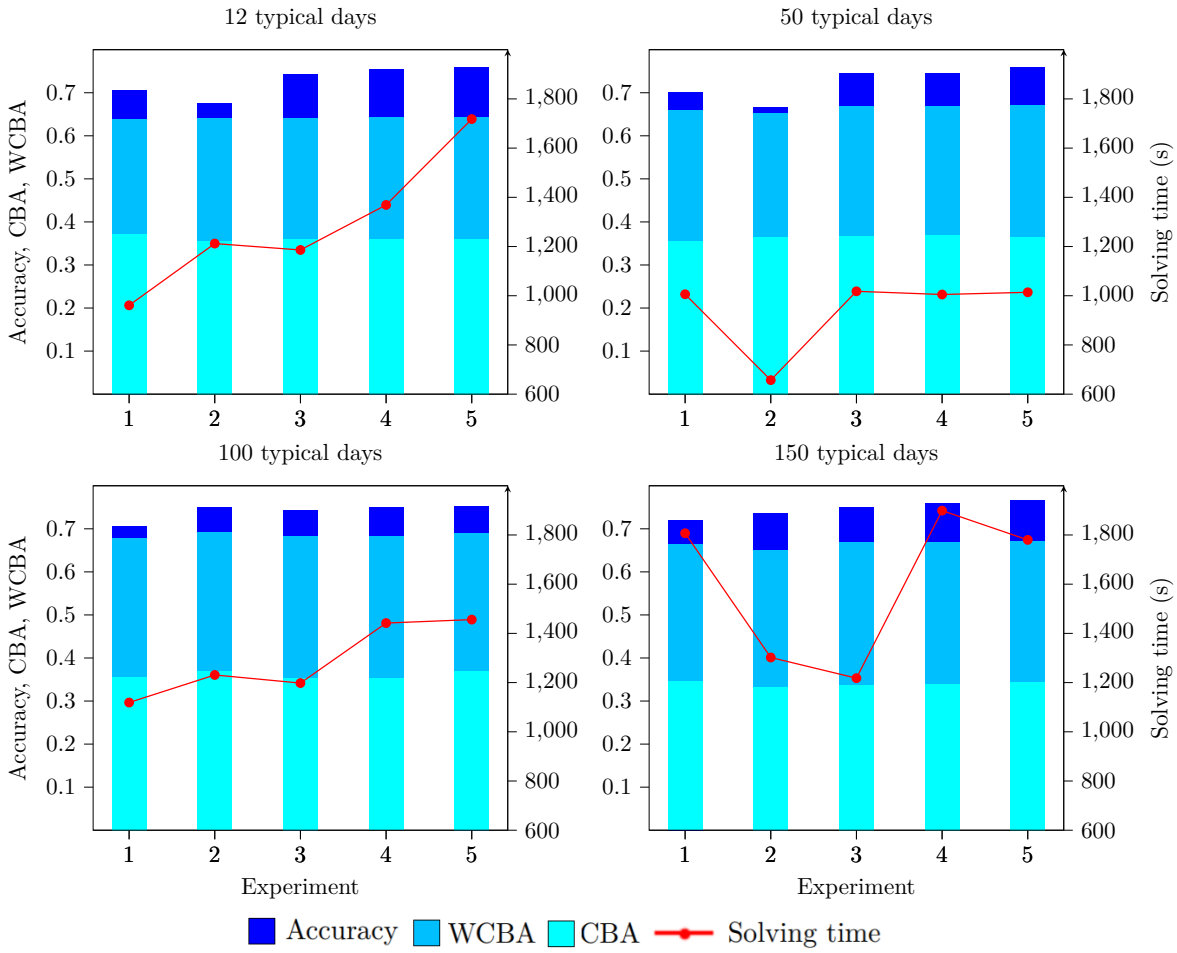

Figure 5.15: Accuracy, WCBA, CBA and solving time obtained for representative experiments using training data generated with different number of typical days for training data of 30 months

The experiments performed with the use of typical days in the operation optimization show similar results than the ones obtained using the original inputs. Apparently, according to figure 5.13 there is almost no difference in using 12 typical days or higher. The accuracy scores range from 0.68 to 0.75 for all of the experiments pictured in figure 5.13. The same occurs for CBA and WCBA, which do not exceed values of 0.37 and 0.7, respectively. The CBA and WCBA scores are slightly lower than the ones obtained for the experiments using original inputs in the training data, while the accuracy scores are almost the same. However,

as was also the case for 18 months, the computation times are significantly lower when using typical days. Although the values of the computation time vary from 600 s to almost 2000 s, the performance of the RF classifier is very similar to the ones pictured in figure 5.13, where the computational times exceeded 2800 s in all cases.

Different experiments have been carried out to evaluate the effects of different conditions on the performance of the RF classifier. The use of typical days, different values for the RF parameters and the use of different training periods have had little effect on the selected KPIs. However, the RF classifier has not been able to return accuracy, CBA and WCBA scores higher than 0.8, 0.4 and 0.75, respectively. All of the experiments have been performed using training data obtained from the operation optimization when using a yearly cycle condition for the charging and discharging cycle of the ice storage. In order to evaluate if this has any effect in the rule extraction process, the operation of the HP will be optimized imposing a daily cycle condition for the ice storage, using the original inputs and also different number of typical days. The results of these optimizations will be used as training data for the rule mining process, in order to evaluate if the daily cycle condition can obtain better results for the rule extraction.

The same way as for the case of the daily cycle condition, the first step is to obtain a benchmark of the best possible performance of the RF classifier. A series of experiments are performed using full year inputs for the training data and high values of RF parameters, the same as described in table 5.1..

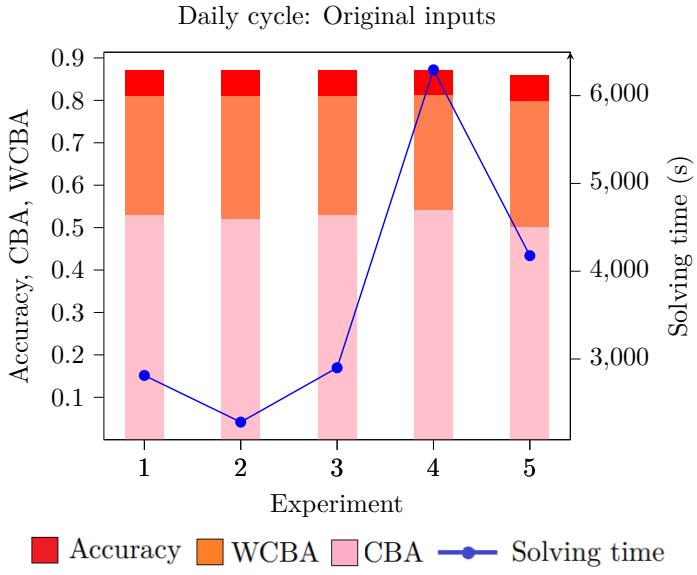

Figure 5.16: Accuracy, WCBA, CBA and solving time obtained for the experiments described in table 5.1 for training data obtained with original inputs and daily cycle constraint for the ice storage

The bench mark for the daily cycle condition shows significantly better results than the one for the yearly cycle. The accuracy scores raise up to almost 90 % for the 5 experiments. However, the greatest increase is seen in the CBA and WCBA results, which exceed, in all cases, 50 % and 80 %, respectively. The solving times are similar than the ones obtained for the yearly cycle bench mark, except the case of experiment n°4, which needed a larger amount of time to reach a solution. There is still a big gap between the results for these two metrics, which means there is still an imbalance in the performance of the RF classifier in the prediction on the different classes. The values of precision and recall of the different classes, for the 5 experiments, are displayed in figure 5.17

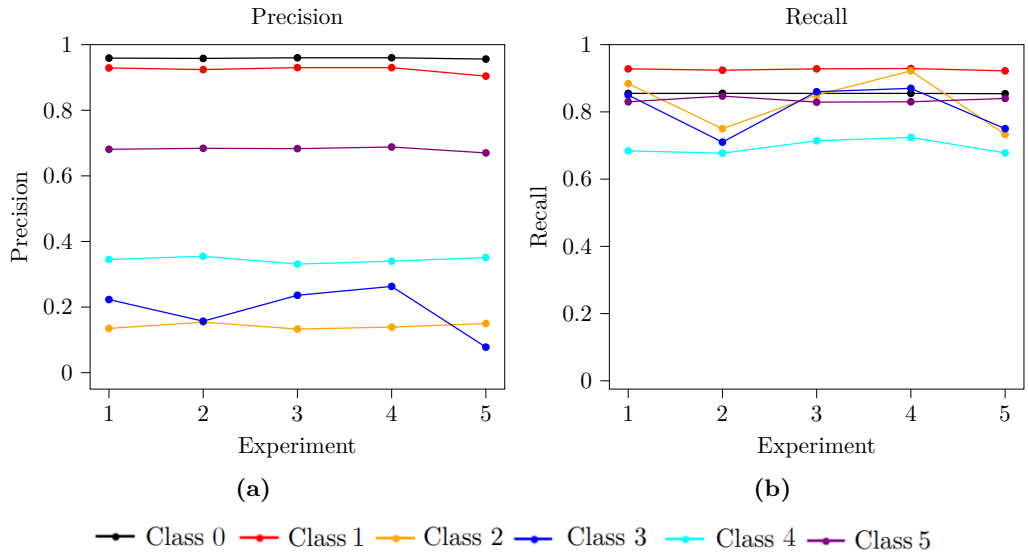

Figure 5.17: Values of precision (a) and recall (b) of each class present in the predictions displayed in figure 5.16

	Cold TRY	Normal TRY	Warm TRY
Off	37%	47.6%	49.5%
Mode 1	41%	36.7%	34.5%
Mode 2	3.7%	2.8%	3%
Mode 3	0.4%	1%	0.4%
Mode 4	8%	5.4%	4.8%
Mode 5	3.9%	6.5%	7.8%

Table 5.4: Relative share of each mode of the HP for an operation optimization with full year inputs and daily cycle condition for the ice storage

There is a significant improvement in the precision and recall scores of the different classes,

compared to the ones when using the yearly cycle condition. The precision of all classes, except class 3, which was suffered an important drop, have increased. The classes with a higher relative share, as class 0 and 1, show almost perfect scores for precision, and the less frequent ones, as class 1 and 2, have undergone a noticeable increase, although still showing low values of precision. However, the greatest difference over the case of the yearly cycle condition can be seen when looking at the recall graph. Except for class 3, the recall score of the rest of the classes exceed 80 %. This is a major improvement over the previous case, although its is not reflected in the CBA score due to the fact that precision scores are lower than the recall scores for almost every class. The increase of CBA and WCBA over the case of the yearly cycle condition is caused by a more accurate prediction of all classes, not just the less frequent ones.

Once the bench mark has been set, the same procedure applied previously will be followed. Several predictions are carried out, for different number of typical days values of the RF parameters.

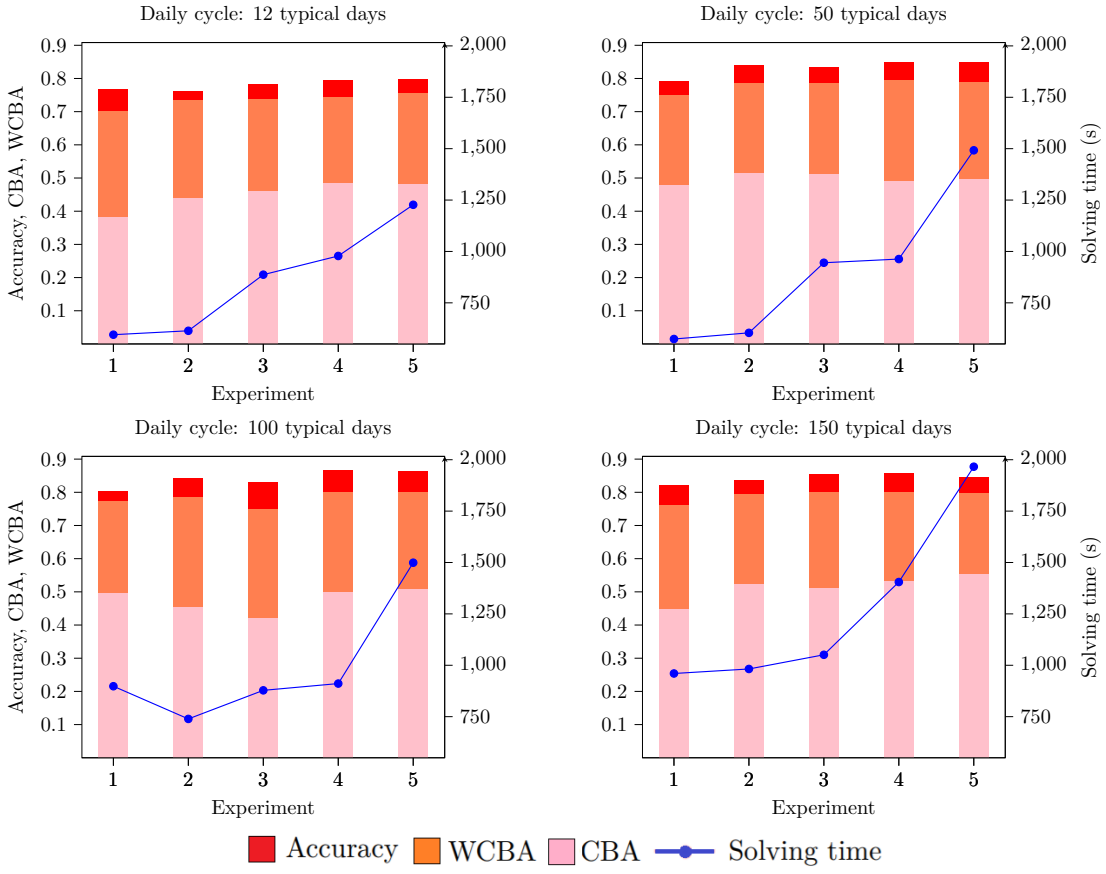

Figure 5.18: Accuracy, WCBA, CBA and solving time obtained for representative experiments using training data generated with different number of typical days and daily cycle constraint for the ice storage

Unlike in the case of the yearly cycle condition, it is possible to observe some differences in the obtained results for different number of typical days. For 12 days, the accuracy, WCBA and CBA scores do not exceed 0.8, 0,75 and 0.5, respectively. The results for the three KPIs slightly increase when using higher number of typical days, reaching maximum values of 0.87 for accuracy, 0.55 for CBA and 0.83 for WCBA.

When using 12 typical days, it is possible to see that the scores for the three KPIs increase, to a small degree, as the solving times get higher. This does not happen when using higher number of typical days, where there is no visible correlation between the accuracy, CBA and WCBA scores and the needed solving time. In any case, figure 5.18 shows similar results when using the original inputs or different number of typical days in the training data. Except for just 1 case, experiment 5 for 150 typical days, all of the experiments used solving times lower than 1500 s. However, for the benchmark, almost all of the experiments exceed 3000 s of solving time.

6 Summary and outlook

Within the scope of this master thesis, the model of the HP system of FUBIC was introduced in an approximate MPC process to extract control rules from it. Inside this process, a k-medoid clustering algorithm was used to cluster the original, full year inputs, into a predefined number of typical days representing these inputs. The aim of the thesis was to evaluate the effects of the use of different number of typical days in the different stages of the approximate MPC process and, especially, in the final extracted control rules.

The first step was to evaluate the performance of the clustering algorithm in the generation of the input curves. The results for the selected KPIs indicate a more accurate representation of the temperature and heating demand input curves than for cooling demand. This was especially noticeable when using 7 typical days or less. One of the reasons for this is the fact that the clustering algorithm is coded with a weight of 2 for temperature and 1 for the rest of the input variables. For this reason, the algorithm is giving a higher importance in the clustering of temperature, and, therefor, generating a more accurate temperature input curve. In future studies, the possibilities of using different weights in the clustering algorithm could be introduced, in order to evaluate if there is a change in the performance of the algorithm.

The clustered curves were used as inputs for the operation optimization of the HP system. The outputs of the optimization were compared, using different number of typical days as well as the original inputs. This was carried out for the cold, normal and warm weather conditions. The results show that, when using typical days, the MILP returned a more imbalanced mode distribution of the optimal operation of the system. This was is especially significant for the cold TRY. It also happens, in a lesser degree, for the normal TRY and it is almost unnoticeable in the case of the cold TRY. The charging and discharging cycles of the energy storages were also an object of study for this thesis. The results suggest that the representation of the charging and discharging cycle of the ice storage is, although being constrained with a yearly cycle condition, better than in the case of the two buffer storages. The last analysed output of the operation optimization is the value of the objective function, which represents the cost of operation of the HP for a period of a year. When using just 12 typical days, the value of the OF suffered a deviation of less than 2.5 % from the value of the OF when using the original, non clustered inputs.

The last part of the analysis consists in evaluating if the use of typical days was a valid approach to extract control rules from the operation optimization. The results show that

the performance of the classifier in predicting the optimal mode schedule for a testing period was very similar when using typical days or original inputs in the training data. The use of two different training periods, 18 and 30 months, does not seem to have a great effect either. Additionally, the use of very high values for the different RF parameters did not improve the performance of the classifier, but it did increase significantly the time needed to obtain a prediction. However, the use of a daily cycle constraint for the ice storage in the operation optimization did have an important effect in the accuracy of the predictions. When using the daily cycle constraint in the training data, the accuracy, CBA and WCBA scores for the predictions suffered a considerable increase. As was the case with the daily cycle condition, the use of typical days in the training data or lower values for the RF parameters had almost no effect in the performance of the classifier.

A notable aspect discovered during the analysis of the rule mining results was the significant difference in the precision and recall scores for the least frequent classes present in the predictions. This improved slightly when changing the training period from 18 to 30 months, and improved again when using the daily cycle constraint for the charging and discharging cycle of the ice storage. However, the precision and recall scores for the least frequent classes are still comparatively low. In order to improve this, it would be possible to include imbalanced classification metrics, as CBA and WCBA, in the training process of the RF classifier. This way, the algorithm would try to obtain predictions with a greater balance between the precision and recall scores of the different classes, and not focus only in the global accuracy of the prediction. Another possibility would be to use a different ML algorithm in the predictions. AddMo allows the use of different classification algorithms, as Artificial Neural Networks (ANN) or Support Vector Machine (SVM), which could have an effect on the predictions carried out in the rule mining process.

Bibliography

[1] *CLARA en r*. https://www.rpubs.com/gregoryq/550380, . – Accessed: 2021-09-15

[2] *Climate emergency declarations in 2,037 jurisdictions and local governments cover 1 billion citizens*. https://climateemergencydeclaration.org/climate-emergency-declarations-cover-15-million-citizens/,

[3] *Clustering*. http://www.pabloruizruiz10.com/resources/Curso-Machine-Learning-Esp/5---Aprendizaje-No-Supervisado/Intro-Clustering.html, . – Accessed: 2021-09-15

[4] *Coefficient Of Performance – What is a Good COP For Heat Pump*. https://www.pickhvac.com/faq/heat-pump-cop/,

[5] *Compression heat pump and chiller*. https://oemof-thermal.readthedocs.io/en/latest/compression_heat_pumps_and_chillers.html

[6] *Energy performance of buildings directive*. https://ec.europa.eu/energy/topics/energy-efficiency/energy-efficient-buildings/energy-performance-buildings-directive$_en$,

[7] *How to normalize the RMSE*. https://www.marinedatascience.co/blog/2019/01/07/normalizing-the-rmse/

[8] *Importance of Distance Metrics in Machine Learning Modelling*. https://towardsdatascience.com/importance-of-distance-metrics-in-machine-learning-modelling-e51395f . – Accessed: 2021-09-21

[9] *Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE)*. http://www.eumetrain.org/data/4/451/english/msg/ver$_cont_var$/uos3/uos3$_k$o1.htm, . – – *Accesed* : 2021 – 12 – 01

[10] Agarwal, Rahul: *The 5 Classification Evaluation metrics every Data Scientist must know*. https://towardsdatascience.com/the-5-classification-evaluation-metrics-you-must-know-aa97784ff226, . – Accessed: 2021-12-01

[11] Alashwal, Hany ; El Halaby, Mohamed ; Crouse, Jacob J. ; Abdalla, Areeg ; Moustafa, Ahmed A.: The Application of Unsupervised Clustering Methods to Alzheimer's Disease. 13, 31. http://dx.doi.org/10.3389/fncom.2019.00031. – DOI 10.3389/fncom.2019.00031. – ISSN 1662–5188

[12] Alok Malik, Bradford T.: *Applied unsupervised learning with R.* Packt, 2019. – 30 S.

[13] Chen, Steven ; Saulnier, Kelsey ; Atanasov, Nikolay ; Lee, Daniel D. ; Kumar, Vijay ; Pappas, George J. ; Morari, Manfred: Approximating Explicit Model Predictive Control Using Constrained Neural Networks. In: *2018 Annual American Control Conference (ACC)*, 2018, S. 1520–1527

[14] Ciglera ; Gyalistrasb, Dimitrios ; Tietd, Vinh-Nghi ; Luká ; Ferkla: Beyond theory : the challenge of implementing Model Predictive Control in buildings Ji ř, 2013

[15] Crespo, Cristina ; O'Rorke, Caitlin: Diez claves del acuerdo que cierra la COP25. In: *National Geographic* (2019), Dec

[16] Domínguez-Muñoz, Fernando ; Cejudo-López, José M. ; Carrillo-Andrés, Antonio ; Gallardo-Salazar, Manuel: Selection of typical demand days for CHP optimization. 43, Nr. 11, 3036–3043. http://dx.doi.org/10.1016/j.enbuild.2011.07.024. – DOI 10.1016/j.enbuild.2011.07.024. – ISSN 0378–7788

[17] Drgoňa, Ján ; Arroyo, Javier ; Figueroa, Iago C. ; Blum, David ; Arendt, Krzysztof ; Kim, Donghun ; Ollé, Enric P. ; Oravec, Juraj ; Wetter, Michael ; Vrabie, Draguna L.: All you need to know about model predictive control for buildings. . – Publisher: Elsevier

[18] Drgoňa, Ján ; Picard, Damien ; Kvasnica, Michal ; Helsen, Lieve: Approximate model predictive building control via machine learning. 218, 199–216. http://dx.doi.org/10.1016/j.apenergy.2018.02.156. – DOI 10.1016/j.apenergy.2018.02.156. – ISSN 0306–2619

[19] Fernando, Jason: *R-Squared.* https://www.investopedia.com/terms/r/r-squared.asp, . – Accessed: 2021-12-01

[20] Germany's National Meteorological Service: *Weather and Climate - Germany's National Meteorological Service - Services - Test reference years (TRY).* https://www.dwd.de/DE/leistungen/testreferenzjahre/testreferenzjahre.html?nn=507312. Version: 2017

[21] Grandini, Margherita ; Bagli, Enrico ; Visani, Giorgio: Metrics for Multi-Class Classification: an Overview. In: *ArXiv* abs/2008.05756 (2020)

[22] Gupta, Akhilesh ; Tatbul, Nesime ; Marcus, Ryan ; Zhou, Shengtian ; Lee, Insup ; Gottschlich, Justin: *Class-Weighted Evaluation Metrics for Imbalanced Data Classification.* 2020

[23] Hofer, Matthias ; Muehlebach, Michael ; D'Andrea, Raffaello: Application of an approximate model predictive control scheme on an unmanned aerial vehicle. In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016, S. 2952–2957

[24] Hoffmann, Maximilian ; Priesmann, Jan ; Nolting, Lars ; Praktiknjo, Aaron ; Kotzur, Leander ; Stolten, Detlef: Typical periods or typical time steps? A multi-model analysis to determine the optimal temporal aggregation for energy system models, 2021

[25] Jiménez, Fernando ; Webster, Mort D.: Optimal Selection of Sample Weeks for Approximating the Net Load in Generation Planning Problems, 2013

[26] Killian, Michaela ; Kozek, Martin: Ten questions concerning model predictive control for energy efficient buildings. In: *Building and Environment* 105 (2016), S. 403–412

[27] Kotzur, Leander ; Markewitz, Peter ; Robinius, Martin ; Stolten, Detlef: Impact of different time series aggregation methods on optimal energy system design. In: *Renewable Energy* 117 (2018), Mar. http://dx.doi.org/10.1016/j.renene.2017.10.017. – DOI 10.1016/j.renene.2017.10.017. – ISSN 0960–1481

[28] Maier, Laura ; Henn, Sarah ; Mehrfeld, Philipp ; Müller, Dirk: Approximate Model Predictive Control for Heat Pumps in Building Energy Systems

[29] Mechleri, Evgenia ; Sarimveis, Haralambos ; Markatos, N.C. ; Papageorgiou, L.G.: Optimal design and operation of distributed energy systems: Application to Greek residential sector. In: *Renewable Energy* 51 (2013), 03, S. 331–342. http://dx.doi.org/10.1016/j.renene.2012.09.009. – DOI 10.1016/j.renene.2012.09.009

[30] Mosley, Lawrence: A balanced approach to the multi-class imbalance problem, 2013

[31] Peruchena, Carlos M. F. ; García-Barberena, Javier ; Guisado, María V. ; Gastón, Martín: A clustering approach for the analysis of solar energy yields: A case study for concentrating solar thermal power plants. In: *Solarpaces 2015: International Conference on Concentrating Solar Power and Chemical Energy Systems* Bd. 1734, 2016 (American Institute of Physics Conference Series), S. 070008

[32] Pinel, Dimitri: *Clustering Methods Assessment for Investment in Zero Emission Neighborhoods Energy System*. 2020

[33] Piotr Plonsky: *Random Forests vs Neural Networks: Which is Better, and when?* https://www.kdnuggets.com/2019/06/random-forest-vs-neural-network.html. Version: 2019

[34] Rao, Krishna: *r^2 or R^2 — When to Use What*. https://towardsdatascience.com/r-when-to-use-what, . – Accessed: 2021-12-01

[35] Remmen, Peter ; Lauster, Moritz ; Mans, Michael ; Fuchs, Marcus ; Osterhage, Tanja ; Müller, Dirk: TEASER: an open tool for urban energy modelling of building stocks. In:

Journal of Building Performance Simulation 11 (2018), Nr. 1, S. 84–98. http://dx.doi.org/10.1080/19401493.2017.1283539. – DOI 10.1080/19401493.2017.1283539. – ISSN 1940–1493

[36] Rätz, Martin ; Javadi, Amir ; Baranski, Marc ; Finkbeiner, Konstantin ; Mueller, Dirk: Automated Data-driven Modeling of Building Energy Systems via Machine Learning Algorithms. 202, S. 109384. http://dx.doi.org/10.1016/j.enbuild.2019.109384. – DOI 10.1016/j.enbuild.2019.109384

[37] Sandels, C. ; Brodén, D. ; Widén, J. ; Nordström, L. ; Andersson, E.: Modeling office building consumer load with a combined physical and behavioral approach: Simulation and validation. 162, 472–485. https://ideas.repec.org/a/eee/appene/v162y2016icp472-485.html. – Publisher: Elsevier

[38] Schütz, Thomas ; Schraven, Markus H. ; Fuchs, Marcus ; Remmen, Peter ; Müller, Dirk: Comparison of clustering algorithms for the selection of typical demand days for energy system synthesis. In: *Renewable Energy* 129 (2018), S. 570–582

[39] Sharma, Pulkit: *4 Types of Distance Metrics in Machine Learning*. https://www.analyticsvidhya.com/blog/2020/02/4-types-of-distance-metrics-in-machine-learning/, . – Accesed: 2021-11-10

[40] Uppada, Santosh K.: Centroid Based Clustering Algorithms-A Clarion Study, 2014

Appendix

Declaration of Originality

I hereby declare that this thesis and the work reported herein was composed by and originated entirely from me. Information derived from the published and unpublished work of others has been acknowledged in the text and references are given in the list of sources. This thesis has not been submitted as exam work in neither the same nor a similar form. I agree that this thesis may be stored in the institutes library and database. This work may also be copied for internal use.

Aachen, Tuesday 14$^{\text{th}}$ December, 2021

Antonio Gil Toral