AI Forecasting for Enhanced Energy Flexibility in Supermarket Refrigeration Systems

Master's thesis in Computer science and engineering

Elias Hallberg
Samuel Widén

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2024

Master's thesis 2024

AI Forecasting for Enhanced Energy Flexibility in Supermarket Refrigeration Systems

Elias Hallberg, Samuel Widén

UNIVERSITY OF GOTHENBURG

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering
Chalmers University of Technology
University of Gothenburg
Gothenburg, Sweden 2024

AI Forecasting for Enhanced Energy Flexibility in Supermarket Refrigeration Systems

Elias Hallberg
Samuel Widén

Supervisor: Milad Malekipirbazari, CSE
Advisor: Gustaf Engström, Ambidex
Examiner: Ashkan Panahi, CSE

Master's Thesis 2024
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Gothenburg, Sweden 2024

AI Forecasting for Enhanced Energy Flexibility in Supermarket Refrigeration Systems

Elias Hallberg, Samuel Widén
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

Due to the rising misalignment between energy supply and demand, energy flexibility is becoming more relevant. Because supermarket refrigeration systems consume a lot of energy, and their innate ability to act as thermal energy storage, they are a prime candidate to utilize for energy flexibility.

This thesis focuses on exploring new parameters which could be used to improve the performance of short term load forecasting models for supermarket refrigeration systems, thereby improving the ability to utilize them for energy flexibility. In addition to this, it compares the performance of multiple different machine learning models.

The thesis demonstrates that parameters that have not previously been utilized for short term load forecasting of supermarket refrigeration systems, such as Google's popular times graph, can successfully be used to improve prediction accuracy.

Keywords: computer science, engineering, machine learning, refrigeration, load forecasting, energy flexibility, AI

Acknowledgements

We are very grateful to our supervisors Milad Malekipirbazari and Gustaf Engström, who have always been there to support us. Thank you!

We would also like to extend our gratitude to our friends who have cheered us on. It would not have been possible without you.

Elias Hallberg, Samuel Widén, Gothenburg, 2024-06-24

Contents

List of Figures

List of Tables

1
Introduction

With renewable energy occupying an ever increasing share of the energy market [1], energy flexibility is becoming more relevant because of the inherent variance in the output of renewable sources. This is made evident by an investigation from Svenska Kraftnät, which concludes that due to the expected increase in wind and solar power, there will be an increased need to align the times of high supply and high demand, through energy storage and other energy flexibility measures [2].

Currently, refrigerated display cases (RDCs) in supermarkets consume approximately 1-1.5% of the total electricity in Sweden [3]. Due to the fact that there is a span of acceptable temperatures in these RDCs, it makes them an ideal candidate for load balancing; the RDCs can be run at maximum effect for a short time if there is an energy surplus in the electrical grid, or can almost entirely halt cooling in response to an energy deficit.

While there exist models that can predict the temperature and load demand of RDCs, they are dependent on accurate input parameters, such as temperature inside the cabinet, temperature in the supermarket, humidity, and frequency of cabinet interactions [4]. This makes them unsuitable for predictions over a longer period of time, such as 24 hours ahead. With models that can accurately predict further into the future, supermarkets would be able to make calculated bids on the energy flexibility market (such as Frequency Containment Reserve (FCR)), which have historically been won by hydroelectric power stations [5]. Since bidding on FCR is done one day ahead, it is particularly important for that use case to have accurate forecasting on this timescale.

With this in mind, an AI-based prediction model for forecasting energy consumption of supermarket refrigeration systems would provide new opportunities for supermarkets to help balance Sweden's electrical grid by engaging with the energy flexibility market, while being monetarily incentivized to do so. This would be an important step towards enabling renewable sources of electricity to make up an even larger share of Sweden's electricity production.

1.1 Goals and Challenges

The goal of this thesis is to create a machine learning model that can predict the required energy consumption of the refrigeration system of a supermarket up to at least 24 hours later. The model will be based on data that is easily available to supermarkets. An important data point that may however be hard to find will be how often the RDCs are interacted with. One way to solve this is to use methods to approximate the popularity of the supermarket at each point in time. There is a multitude of ways to find such an approximation, for example Google's popular times metric, and most supermarkets even have a count of the number of people in the store at any given moment.

One of the goals is that the model should be usable for most supermarkets in the world as long as it is trained on data from the store. Thus, the chosen input parameters should capture most of the causality for the energy expenditure of the refrigeration system. On the other hand, if any of the assumptions that are made do not apply to a certain store, the model would lose some level of accuracy. An example of this would be if the humidity in a store was to vary greatly, especially within one season.

The primary challenge is the fact that it is very difficult to accurately predict all of the required parameters a day ahead. This is why an AI system is ideal, as it can utilize data from many different sources to create a better prediction. This is also why the parameters have been limited to the ones believed to be the most important that are also easy for the supermarkets to gather.

One other challenge is that while the temperature inside the RDCs can be monitored, it is the temperature of the products that matters in terms of food safety. This is a problem because it is much more difficult to measure their temperature. The difference in temperature of the RDCs and food products is not an issue during normal operation of the RDCs, as the temperature inside is on average equal to the temperature setpoint, which means the temperature of the products will also equal this setpoint. However, if the RDCs were to be shut off temporarily, such as when there is an energy deficit, the temperature of the air will rise faster than the temperature of the products. This means that in order to fully utilize the RDCs for demand response, models for the temperature inside the food will necessary.

As the energy expenditure of the supermarket is likely to correlate with the season, it could theoretically be a problem that there is only one year of data available from the supermarket. This is because it would be difficult to train a model for seasonal trends if the season has only been observed once. However, since the underlying reasons for the difference in energy expenditure by season are captured by the input parameters, the model should be able to handle any season regardless.

1.2 Limitations and Scope

While there are many different frequency containment reserves in Sweden and worldwide, the focus of this paper will be on Svenska Kraftnät's Upwards Frequency Containment Reserve - Disturbance (FCR-D). The reason for this is that the data is from a Swedish supermarket and thus the Swedish frequency containment reserves are the most relevant. The reason that FCR-D is chosen is because it has lower endurance requirements for limited energy resources (LERs), which is important because there are limits to how long the RDCs can be turned off before the food gets too hot.

Based on the endurance constraint for FCR-D, in order for supermarkets to bid, it must be feasible to turn off the refrigeration of the RDCs for 20 consecutive minutes. This in turn requires that the food in the RDCs will not become too hot during this period. Due to the fact that the RDCs and the food inside will take longer to warm up than the air in the RDCs, we will initially assume that turning off the refrigeration for 20 minutes does not violate the temperature constraints for any food in the RDCs. As far as we know, this assumption is also in line with the standard industry praxes.

1.3 Risks and Ethical Concerns

One potential ethical concern is that the system, if used to facilitate bidding on the energy flexibility market, will incentivize shutting off the refrigeration of the RDCs, which could possibly affect the food in negative ways and could pose a health hazard. Therefore, the completed thesis must clearly state what assumptions are made, and what guarantees can be given in terms of food safety. Of course, the goal is to arrive at a system that is known to be safe for its intended use case, but because the confirmation of such safety is possibly outside the scope of the project, then the potential risks must be conveyed. Still, in accordance with EU law, the ultimate responsibility when it comes to making sure the food is safe for consumption is on the producers and the supermarkets, so it is reasonable to assume they will take appropriate measures as long as they are informed of the potential hazards.

Because data from a supermarket will be used, one may have concerns about privacy issues. However, no individually identifiable or sensitive data will be used, and thus there is no possibility of encroaching on privacy.

2

Background

In this section, the various factors affecting the energy consumption of refrigerated display cases in supermarkets will be explored. Previous research into the subject of load forecasting for RDCs will also be presented in this section. Furthermore, the data that was used will be described.

2.1 Context

While supermarket refrigeration systems are well suited for load balancing, the time that they can increase or decrease energy consumption is still relatively short compared to other resources. This makes them fall under the category of LERs. The energy flexibility market that is most relevant for LERs is the FCR-D, as the requirements on endurance are drastically lower than for other markets. Recently, Svenska Kraftnät concluded a pilot study into how LERs can contribute to the energy flexibility market. This resulted in the lifting of some of the restrictions that had previously been applied to LERs that were bidding on or wanted to bid on the energy flexibility market [6]. Therefore, the incentives for supermarkets to bid on FCR-D were increased, and it is more relevant than ever to develop a model that enables them to do so.

The bidding on FCR-D is done in two auctions per day. The first auction closes at 00:30 CET the day before, and the second auction closes at 18:00 CET the day before. As such, predictions ideally need to be made at least 30 hours ahead.

Supermarket refrigeration systems typically utilize Vapor-Compression Refrigeration (VCR). In VCR systems, gaseous refrigerant that has been heated by the RDCs enters a *compressor* and gets compressed to a higher pressure, which raises the temperature. The hot refrigerant then enters a *condenser* (located outside the store, typically on the roof), where it is cooled and condenses. The condensed liquid refrigerant then passes through an *expansion valve*, where it undergoes an abrupt reduction in pressure. This results in the evaporation of a part of the liquid refrigerant, which lowers the temperature of the liquid and vapor refrigerant mixture such that it is colder than the temperature of the RDC. Finally, the cold refrigerant mixture passes through the *evaporator* (located by the RDC), where heat is transferred from the RDC to the refrigerant. This causes the refrigerant to completely vaporize, after which it is once again routed into the compressor, and the cycle is complete. Because the condenser is located outside, the outside temperature has an effect on the power consumption of the refrigeration system. Specifically, the work that the compressor has to do when compressing the refrigerant is increased, as a higher outside temperature requires a higher temperature of the refrigerant in the condenser in order to achieve the same amount of cooling, in turn necessitating a higher pressure.

As supermarkets are responsible for ensuring the food they sell is safe for consumption, they must consider the food safety guidelines outlined by Livsmedelsverket [7]. The guidelines contain the recommended maximum temperatures for each different type of food, however, this of course refers to the temperature of the food, rather than the temperature of the air in the RDC. As such, assuming that the temperature in the RDC is normally within the recommended range, there is an opportunity to allow the temperature of the air in the RDC to temporarily rise above the recommended maximum, as the food and the metal in the RDC itself will only slowly start to begin to heat up from the warmer air.

There exists prior research into utilizing supermarket RDCs for energy flexibility, such as the hygro-thermal model by Månsson et al. [4]. However, while the model is accurate, it does not predict far enough into the future to be useful for energy flexibility bidding. This seems to be the trend for the field in general, with other models also achieving a relatively good result, but only for minutes ahead. Trying to predict the energy consumption longer into the future is hard because there are many different parameters that affect it, and it is difficult to predict how they will change over time. We believe that multi-modal AI will be especially beneficial in this area, enabling more accurate predictions over time by utilizing more data.

An attempt to predict over a longer period of time was made by Rasmussen et al. [8], who presents a grey-box model that forecasts the load up to 42 hours ahead. However, besides power, the model only utilizes the outside temperature and whether the store is open or not as input parameters. A system that utilizes more parameters as well as a different model, such as a neural network, is therefore likely to generate a more accurate forecast.

This shows that there are still significant advancements to be made in extending how long the energy consumption can be accurately predicted, and that there is also a need for the models to consider multiple parameters in order to be able to be used effectively.

2.2 Factors Affecting the Energy Consumption of Refrigerated Display Cabinets

The amount of energy expended by the RDCs depends on the heat transfer rate, as well as the efficiency of the compressor, since the heat transfer rate determines how much heat needs to be removed from the system, and the efficiency of the compressor determines how much energy is required to remove that heat. While the heat transfer rate of an RDC depends on the type, by far the most common type of RDC is the vertical closed display cabinet. For such an RDC, the factors affecting the heat balance are mostly related to the infiltration of ambient air, sensible heat gains from transmission through the envelope and glass doors, residue heat from mechanical work by the fans (as well as the lights), and heat transfer by radiation from the ambient.

For the heat transfer from air infiltration, the relevant factors are the amount of air infiltration, the temperature of the ambient air, and the humidity of the ambient air (the humidity affects the enthalpy of the air). The amount of air that infiltrates the RDC can be divided into passive infiltration and active infiltration. Passive air infiltration occurs constantly, due to gaps between the glass doors of the RDC as well as holes in the glass that act as handles for opening the door. Active air infiltration occurs as a result of cabinet interactions (door openings). The temperature of the infiltrating air can be assumed to be relatively constant, as the HVAC system of the supermarket attempts to keep a constant pleasurable indoor temperature, and likewise for the humidity. As all other factors are mostly constant, the greatest variable affecting the heat extraction rate from air infiltration is variations in the frequency and characteristics of door openings, as concluded by Månsson et al. [9].

The other sources affecting the heat balance are either effectively constant or negligible. The fans and lights generate a constant amount of heat (with the exception of the lights being turned off while the store is closed), and a study by Faramarzi et al. [10] found that the heat transfer from radiation only contributes 1.3% of the total thermal load.

Beside the heat balance, another factor that influences the energy consumption of RDCs is the outdoor temperature. This occurs because the condenser, where the refrigerant is cooled, is located outdoors. This means that if the outdoor temperature is very high, the compressor must exert more work to compress the refrigerant to a higher pressure so that the temperature difference between the refrigerant and the outdoor air is sufficient. Therefore, cooler outdoor temperatures means less compressor work, and thus less energy expenditure, and the reverse is true for higher outdoor temperatures.

2.3 Approach

The first step was to perform a literature review of the current methods for estimating energy expenditure of RDCs, as well as the latest approaches for modelling AI systems based on time-series data.

In terms of constructing the model, first the required data was gathered. Some of the data that is of interest is the temperature inside and outside the supermarket, how much the RDC is interacted with, and when the supermarket is open. Most of this data was provided by the supermarket. This data need not be processed very much, except for possible considerations of effect of the compressors as a fraction of the maximum effect, as opposed to the effect in Watt. The data that is missing is how often the RDC doors are opened, which is hard to determine and the supermarket does not have preexisting data for it. To solve this, the busyness of the supermarket was approximated. This was achieved by using popularity values from Google.

When the data had been gathered, the next step was to decide which machine learning models to use. Some models that were considered are ARIMA, convolutional neural networks, and boosted random forest, due to their applicability in this subject or popularity in the industry. The decision on which models to implement was finalized after the literature review. The models were trained on past data and will not be updated continuously. When the types of models were determined, the next step was to create and optimize them. Finally, the models were tested and compared on data from the supermarket.

2.4 Literature Review

There are many different methods that can be used for electric load forecasting, but because models are trained on different datasets it is difficult to compare their performance. To remedy this, Gasparin et al. [11] looked into which models are typically used for load forecasting, and then compared their performance for five different use cases based on three datasets. Use case 1, 2, and 3 are all individual households, which brings very challenging and noisy dynamics. Use cases 4 and 5 are both aggregated load, meaning the load profiles are much more regular. The authors also explain how the models work, and go into detail about the five different architecture-independent strategies for multi-step-ahead forecasting, the two most relevant of which are Recursive and Multiple input-Multiple output (MIMO). Under the recursive strategy, a model is trained to perform a one-step-ahead forecast given some input sequence. The output is then recursively fed back into the model to create the next forecast, and this is repeated as many times as the number of timesteps to forecast for. In the MIMO strategy, a model is trained to forecast the whole output sequence in one shot, meaning the output is a vector rather than a scalar.

From the results, it is clear that there is no single model or strategy which is the best for all scenarios, as the performance of the models vary greatly based on the use case. Still, the authors note that often, the simpler versions of the models perform comparably to the computationally more intensive ones, which would provide a more cost effective solution. The authors also emphasize that encoder-decoder models do not represent a golden standard in load forecasting, unlike in other domains.

While only a tiny fraction of the literature on short term load forecasting focuses specifically on load forecasting of supermarket refrigeration systems, Rasmussen et al. [8] presents time adaptive linear time series models for that purpose. The data was collected from a supermarket located in a village in Denmark, and while the sales area is not mentioned in the article, the energy consumption shows that it is likely a small supermarket. The data used for the forecast models consists of hourly load measurements, local measured ambient temperature and numerical weather predictions for a period of 3 months (May, June, July). Every hour, the models forecast the hourly load for refrigeration for the following 42 hours. The dynamic relations between the inputs and the load are modeled using linear transfer functions and non-linearities are handled with spline functions. The models are fitted using a recursive least squares scheme. The models also operate under two different regimes, one during opening hours and another during closing hours. Notably, it was found that the interval that best separates the two regimes is not in fact the opening and closing hours, but rather 1 hour prior to opening until 1 hour after closing, 7 am to 10 pm. While the outdoor ambient temperature is expected to affect the energy consumption of the supermarket, it was found that this relationship is nonlinear. Furthermore, because the refrigeration system is also interacting thermally with the building, a low-pass filter was introduced, however it was found that the low-pass filtering did not have a big influence on the forecasts. The models proved a significant improvement over the naive method of taking the energy consumption from the same time last day, and based on an analysis of the residuals after applying an auto-regressive noise model the authors argue that only little further improvement to the one-step ahead residuals is possible.

While the models seem very promising, there are some areas where we believe improvements could be made. As mentioned, the models operate in two different regimes, one for opening hours and another when the store is closed. It makes intuitive sense that the energy consumption of the store would be different when open or closed, and it can be seen in the results that it was indeed beneficial to make this distinction. However, we believe that one of the primary reasons for the difference in energy consumption is that the number of cabinet interactions is zero when the store is closed, thus lowering how much heat enters the RDCs, thereby lowering the amount of work the compressors need to exert. Instead of only considering whether the store is open or closed, we believe that the number of cabinet interactions is an even better indicator. Furthermore, the outside temperature is fed to the models, which is beneficial as it positively correlates with how much work the compressors have to do to sufficiently heat the refrigerant. However, the temperature inside the store is possibly even more important, as it affects the heat balance of the RDCs, thereby correlating with the needed energy expenditure. It is not unreasonable to assume that the inside temperature is constant, but it would be wise to check that this is indeed the case, because if it is not it could have a large impact on the overall energy consumption of the refrigeration system.

2.5 Data

To be able to utilize machine learning, data is needed. The data that was gathered for potential use by the models was the popularity of the store, the temperature outside and inside the store, the temperature and moisture inside the RDCs, utilization of the compressors and energy usage for the refrigeration system. As the data was gathered from different sources, they have different sampling frequencies.

2.5.1 Utilization and Effect of Compressors

Utilization of compressors is in percentage of the compressors' max usage and given by the supermarket. The goal was that this data could be used to find the energy usage for the refrigeration systems as while data for energy usage of the refrigeration system was procured from the supermarket, they had only started gathering it recently, while they had data for utilization from far earlier. The hypothesis was that compressor utilization and energy expenditure would have a linear relation, but this turned out to be false, so at this time the utilization of compressors is not usable. Figure 2.1 shows that there indeed seems to be a linear relationship between compressor usage and effect, but there is also a lot of noise. The energy usage for the refrigeration systems that is given by the supermarket is in Watt. The data for the utilization exists for almost a year while the data for the energy usage for the refrigeration systems is from the end of December to the end of april. The sampling frequency for the energy usage is very high, roughly once per second. See figure 2.2 for an example of how the power consumption varies during a week.

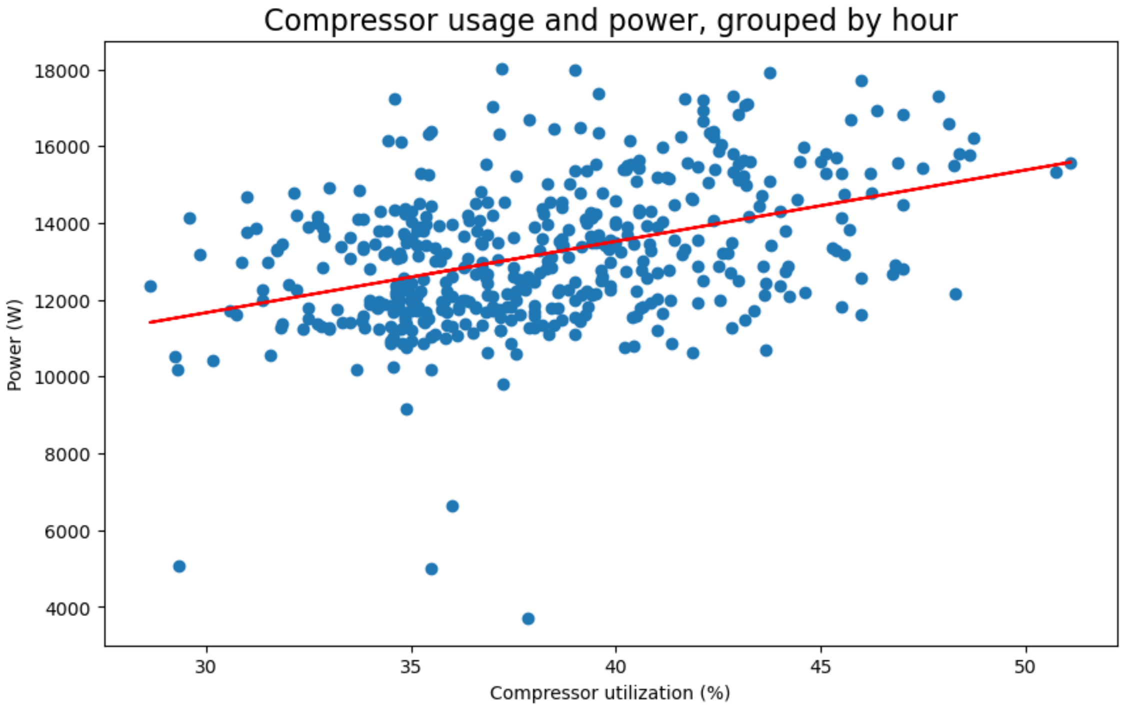

Figure 2.1: Compressor usage and power, grouped by hour

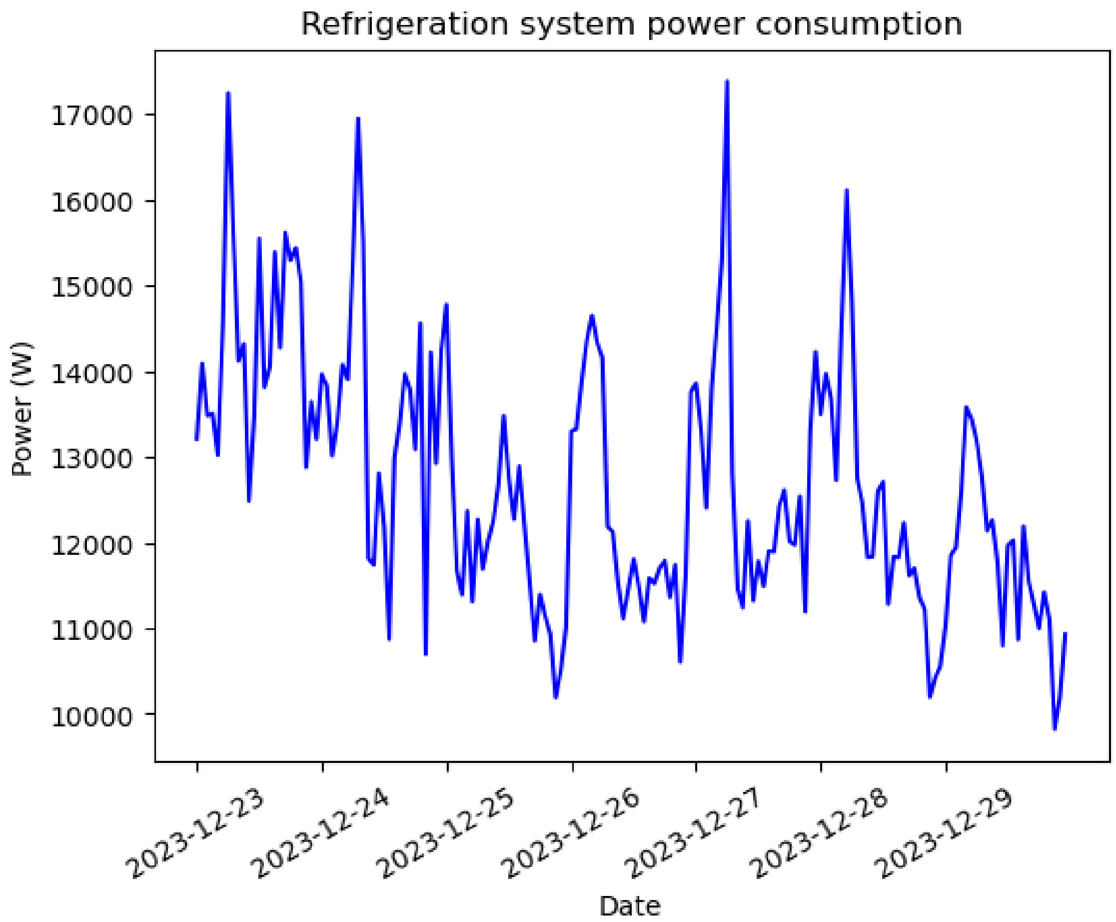

Figure 2.2: Energy consumption of the refrigeration system for a period around Christmas

2.5.2 Popularity

The popularity of the store was gathered from Google's popular times graph. This data is important because it is used as an approximation of RDC interactions, which influences the heat balance of the RDC. The predictions for the day after are gathered once every day. The actual popularity of the store has then been gathered for different times and days and compared to make sure that the predictions were good. Since the store has been reluctant in sharing data for the number of visitors, we are unable to confirm if Google's popularity correlates with the actual number of people in the store. The data is from the end of February to May. In Figure 2.3 the predicted popularity and the actual popularity is shown for a couple days to showcase the accuracy of the predictions. The predictions are not particularly accurate, but they at least trend in the correct directions. Because there is currently no reliable approximation for RDC interaction beyond a few months, that is the bottleneck for how much data the machine learning models will be able to utilize.

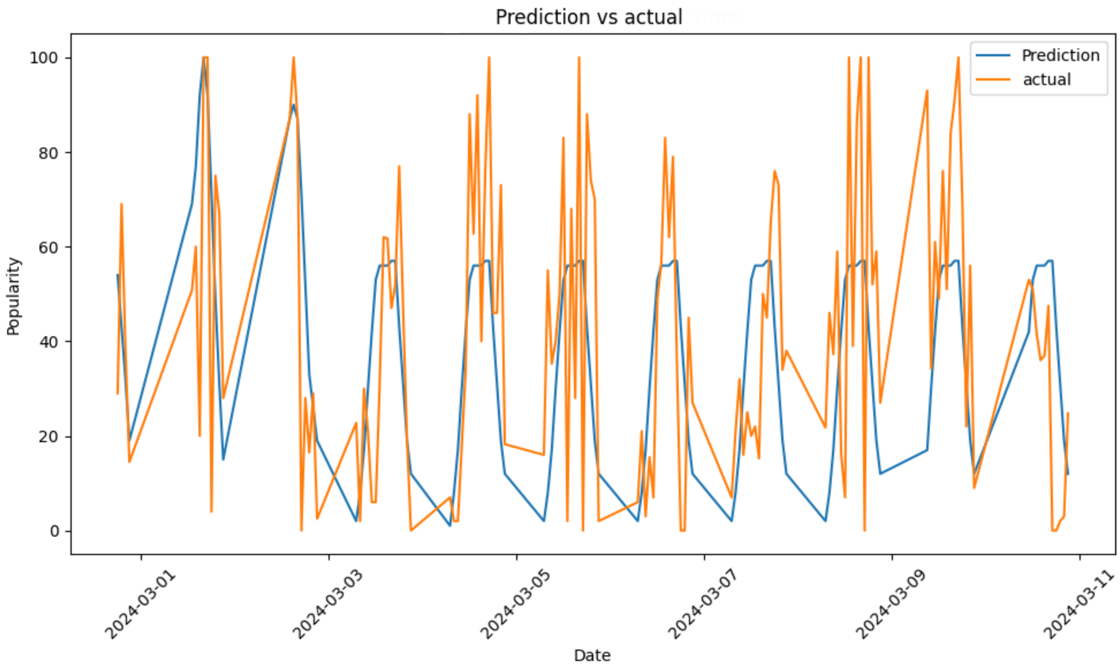

Figure 2.3: Correlation between predicted and actual popularity

2.5.3 Temperature Outside the Supermarket

The temperature outside the store was gathered from Open-Meteo's historical weather API [12] which has temperature as well as other potentially useful data such as relative humidity. The sampling frequency is once per hour. Figure 2.4 shows the temperature outside the store over two months. Data exists for a very long time back.

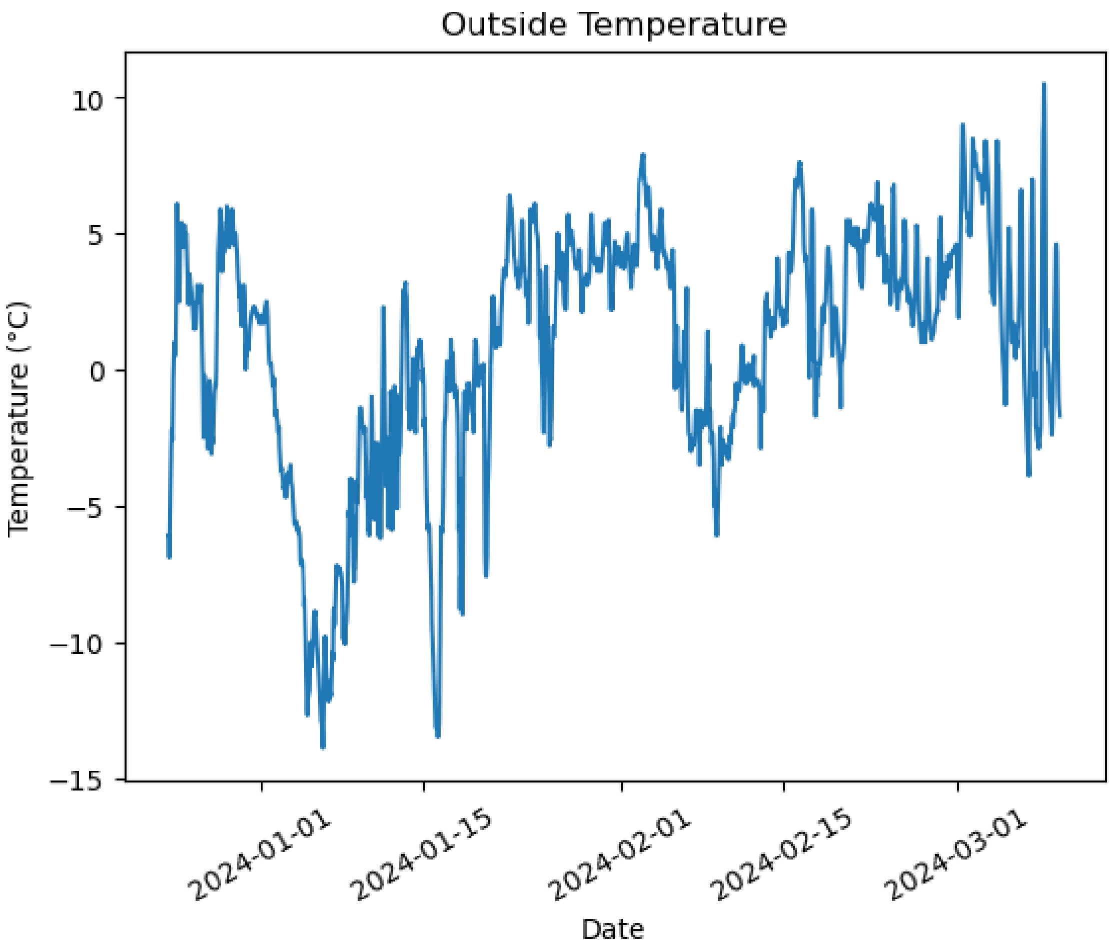

Figure 2.4: Outside temperature

2.5.4 Temperature inside the RDCs

The temperatures inside the RDCs were given by the supermarket and has a sampling frequency of multiple times per hour. There is data for multiple different refrigerators and freezers, which all have different temperatures. Figure 2.5 shows the temperature for one freezer from the supermarket, where the effects of the defrosting is very prevalent. There exist data for about one year.

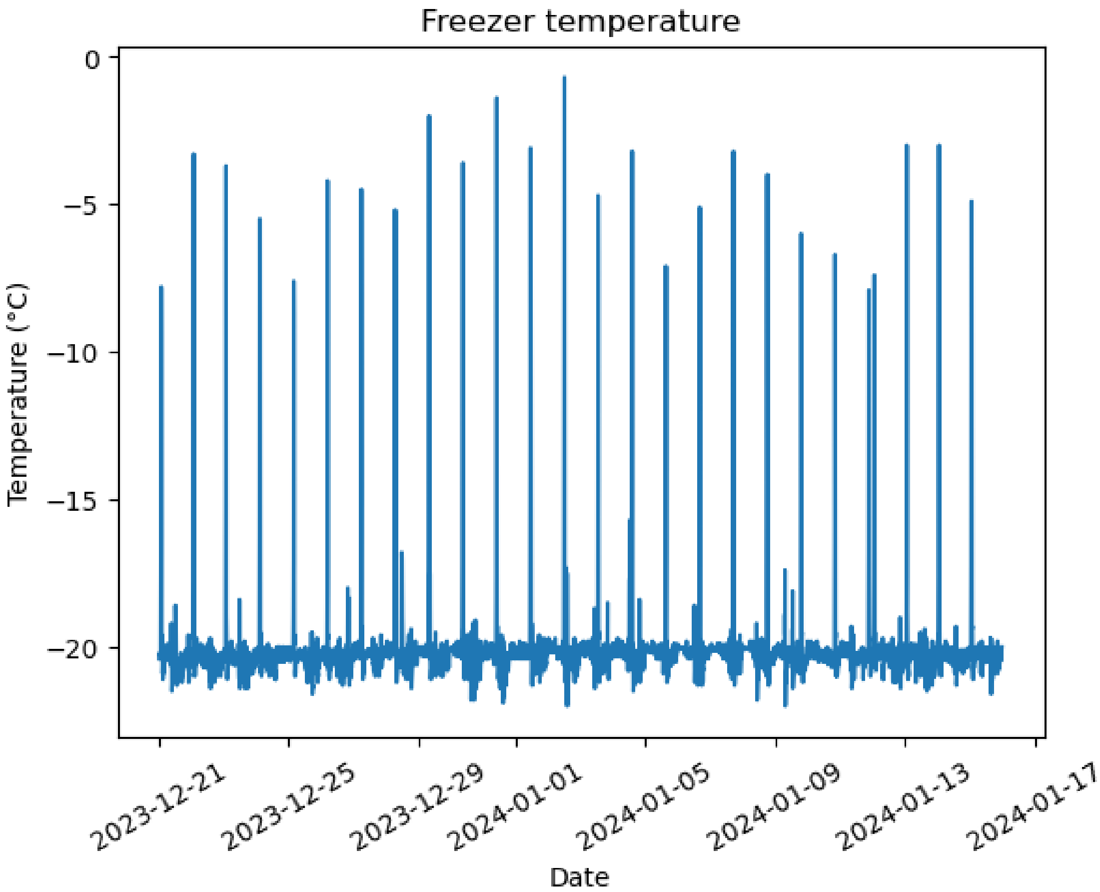

Figure 2.5: Temperature data for one of the freezers

2.5.5 Temperature Inside the Store

The data for the temperature inside the store was acquired from the supermarket. The temperature is measured at a few different locations inside the supermarket, with a sampling frequency of six times per hour. However, an unknown error with the measuring equipment caused the temperature readings to be unusable for the majority of the dataset. Still, from the part of the data that is not corrupted it can be concluded that the temperature inside is relatively stable, and thus not being able to utilize it should not have a negative impact on the performance. Figure 2.6 below shows a representative example of how the temperature varies over the day.

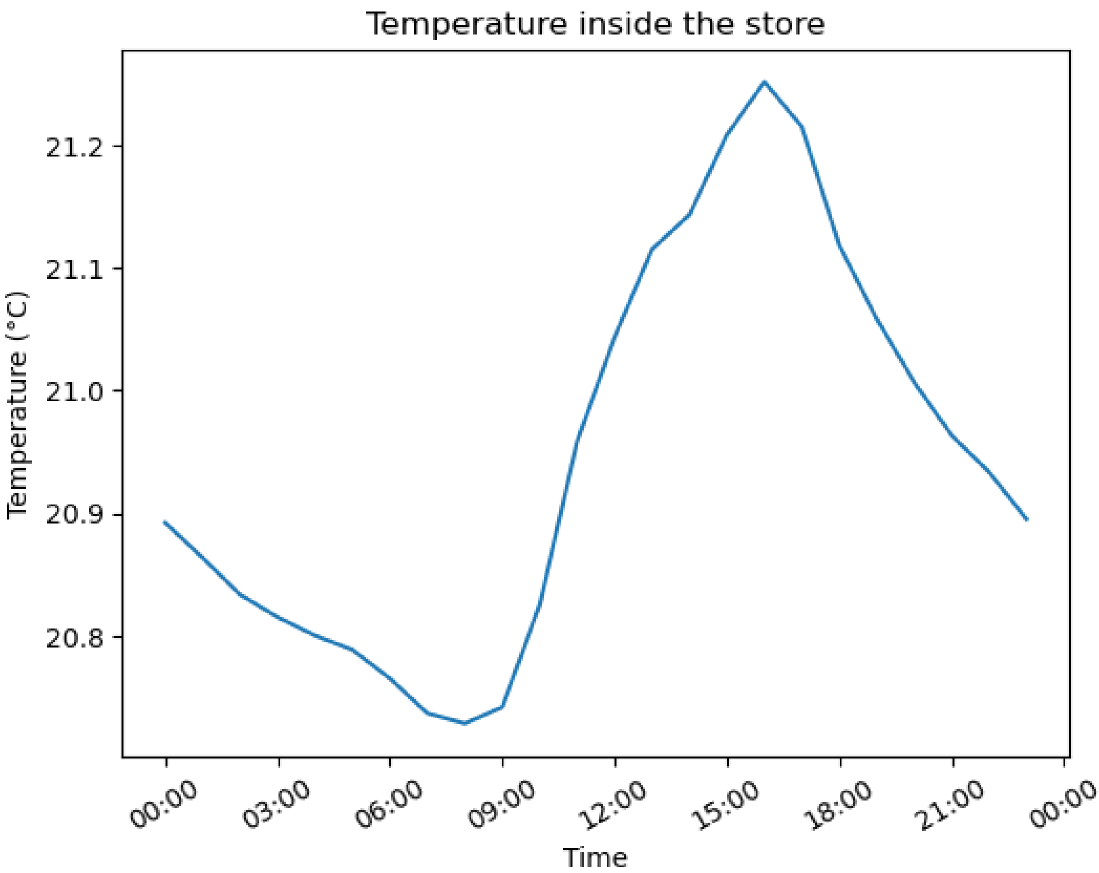

Figure 2.6: Temperature inside the store

3

Methods

3.1 Models

This section explains the history and theory of the models that were used in the thesis. The models that were used were chosen because they have been used in similar projects and have all had good results for various time series datasets. The models have been trained on approximately 3 months of data.

3.1.1 Temporal Convolutional Network - TCN

Temporal Convolutional Networks (TCNs) were first proposed in 2016, where it was used in an encoder-decoder manner for action segmentation of videos, and showed performance on par with or exceeding traditional models such as LSTM [13]. Since then, further enhancements have been made to the concept. This section will describe the crucial aspects of a TCN, based on our experience and the evaluation by Bai et al. [14].

As implied by their name, TCNs are a type of convolutional network, but they specifically utilize full convolution, to ensure that the length of the output sequence is the same as the length of the input. Note also that TCNs use 1D convolutions, rather than the 2D convolutions typically used by CNNs for image processing. Furthermore, TCNs ensure that there is no data leakage from the future into the past, by employing causal convolutions. A causal convolution is the same as a normal convolution, except that the output at time t is convolved only with elements from time t and earlier in the previous layer.

An issue caused by causal convolution is that the network can only look back at a history with size linear in the depth of the network. To solve this, TCNs use dilated convolutions, which allows the receptive field to grow exponentially. Dilation can be explained simply as introducing a fixed step between each filter tap. The size of this step is called dilation factor, and a dilation factor of 1 would be the same as a normal convolution.

One final common technique used by TCNs is residual connections. Residuals first saw widespread use in image recognition, where it was shown to significantly improve the performance of deep networks [15]. For a residual block, instead of simply passing an input through a number of layers, the input to the block is saved and added to the output, before finally passing it on to the next block. This enables layers to learn modifications to the identity mapping instead of the entire transformation.

3.1.2 Recurrent Neural Networks - RNN

Recurrent Neural Networks (RNNs) were first conceptualized by David Rumelhart in 1986 with the introduction of backpropagation to networks possessing recurrent connections, inspired by earlier architectures like Hopfield networks.

RNNs are widely employed in modeling time series data due to their ability to capture sequential dependencies. Unlike traditional feedforward neural networks, RNNs possess connections that form directed cycles, allowing them to retain information from previous inputs through a mechanism known as Backpropagation Through Time (BPTT) [16]. This enables RNNs to maintain an internal state, crucial for when sequences is processed in AI models.

At each time step t, an RNN maintains a hidden state h_t, which encapsulates information from preceding elements in the sequence. This hidden state is updated based on the current input x_t and the previous hidden state h_{t-1}, following the equation:

$$h_t = f(W_h \cdot h_{t-1} + W_x \cdot x_t + b)$$

where W_h, W_x, and b are the weight matrix for the previous hidden state, the weight matrix for the current input, and the bias vector, respectively. f denotes a non-linear activation function such as sigmoid or hyperbolic tangent (tanh), responsible for updating the hidden state.

Despite their effectiveness, RNNs often encounter challenges such as vanishing or exploding gradients during training, especially with lengthy sequences. This phenomenon leads to issues like overfitting. Techniques such as gradient clipping and specialized RNN architectures like Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs) have been developed to deal with this and improve the performance of RNNs.

3.1.3 Long Short-Term Memory - LSTM

Long Short-Term Memory networks were introduced in 1997 and were created as a specialized version of RNNs designed to mitigate the vanishing gradient problem encountered in traditional RNNs, particularly when dealing with long-range sequential data [11]. LSTMs achieve this by introducing gated mechanisms that regulate the flow of information through the network, allowing it to retain relevant information over long sequences while discarding irrelevant details.

An LSTM cell consists of several gates, including a forget gate, an input gate, and an output gate, which control the flow of information. The forget gate determines what information from the previous cell state should be discarded, while the input gate decides which information from the current input and the previous cell state should be used to update the new cell state. The output gate determines what information from the current cell state to send to the hidden state.

The equations for an LSTM cell are as follows:

$$
\begin{aligned}
i_t &= \sigma(W_i \cdot x_t + U_i \cdot h_{t-1} + b_i) \\
f_t &= \sigma(W_f \cdot x_t + U_f \cdot h_{t-1} + b_f) \\
o_t &= \sigma(W_o \cdot x_t + U_o \cdot h_{t-1} + b_o) \\
c_t &= f_t * c_{t-1} + i_t * \tanh(W_c \cdot x_t + U_c \cdot h_{t-1} + b_c) \\
h_t &= o_t * \tanh(c_t)
\end{aligned}
$$

In these equations, x_t is the input at time step t, h_{t-1} is the hidden state from the previous time step, i_t, f_t, and o_t are the input, forget, and output gates respectively (computed using weight matrices $W_i, W_f, W_o U_i, U_f, U_o$, along with biases b_i, b_f, b_o). c_t is the current cell state, which is a combination of the previous cell state c_{t-1} and the candidate cell state$(\tanh(W_c \cdot x_t + U_c \cdot h_{t-1} + b_c)).h_t$ is the new hidden state. The activation function σ is almost always the sigmoid function, but tanh can be any non-linear activation function depending on the application.

The disadvantage that LSTMs have compared to traditional RNNs is their increased complexity, which often results in higher computational costs, longer training times, and a steeper learning curve for implementation and understanding.

3.1.4 Gated Recurrent Unit - GRU

Gated Recurrent Unit were introduced in 2014 as a variant of RNN which uses gates to handle the vanishing gradient problem without needing to be as complex as LSTMs [11]. It simplified LSTM networks by combining the forget and input gates into a single gate. The benefit of this is a more resource sparse model which leads to it being faster to train while often not having worse results then LSTM models.

The equations for a GRU cell are as follows:

$$
\begin{aligned}
z_t &= \sigma(W_z \cdot x_t + U_z \cdot h_{t-1} + b_z) \\
r_t &= \sigma(W_r \cdot x_t + U_r \cdot h_{t-1} + b_r) \\
\tilde{h}_t &= \tanh(W_h \cdot x_t + U_h * (r_t \cdot h_{t-1}) + b_h) \\
h_t &= (1 - z_t) * \tilde{h}_t + z_t * h_{t-1}
\end{aligned}
$$

In these equations, x_t is the input at time step t, h_{t-1} is the hidden state from the previous time step, z_t and r_t are the update and reset gates respectively (computed using weight matrices W_z, U_z and W_r, U_r, and biases b_z and b_r), $\tilde{h}_t$ is the candidate hidden state (computed using the weight matrices W_h, U_h, and bias b_h), and h_t is the new hidden state. The activation function σ is almost always the sigmoid function, but tanh can be any non-linear activation function depending on the application.

The disadvantage that GRUs have compared to traditional RNNs is the same as LSTM but just as their advantages are less pronounced so are their disadvantages so they are in the middle between RNN and LSTM regarding complexity and ity to handle long-term dependencies.

4

Results

In this section we discuss the model design, and compare the performance of the models for different input features and parameters.

4.1 Model Design

The features used by the models and the design of the models are described in this section.

4.1.1 Input features

The features used by the models are Power consumption, outside temperature, and predicted busyness or observed busyness (from Google's popular times). The reason for using either predicted or observed busyness is to compare the maximum possible performance that could be attained by using Google's popularity metric if the predictions were perfectly accurate, to the performance when using the predictions in their current state.

4.1.2 TCN

The TCN model used is relatively simple, and consists of a small number of residual blocks. Each residual block consists of five parts. First, a 1D convolution layer, with an exponential dilation rate, meaning each subsequent block has twice the dilation rate of the previous block. Then, an activation layer to introduce non-linearity, followed by a dropout layer to prevent overfitting. Fourth, the residual; a 1x1 convolution is applied to the input of the block, to convert it to the same dimensions as the output from the dropout layer. Finally, the result from the dropout layer and the residual are added together, and returned as the output of the block. Figure 4.1 shows the structure of the residual blocks.

The model differs in one major way from other TCNs. Typically, TCNs use causal convolutions, meaning that the output at time t is convolved only with elements from time t and earlier in the previous layer. However, our model uses normal convolutions instead of causal convolutions. This is because the data fed to the model is already time shifted, such that all data points in the output occur at least 24 hours after the final data point in the input. Thus, even without causal convolutions there is no data leakage from the future into the past.

For hyperparameter optimization, the number of convolution filters and the kernel size were chosen simply by trying out a few different values and seeing what performed well on the validation set.

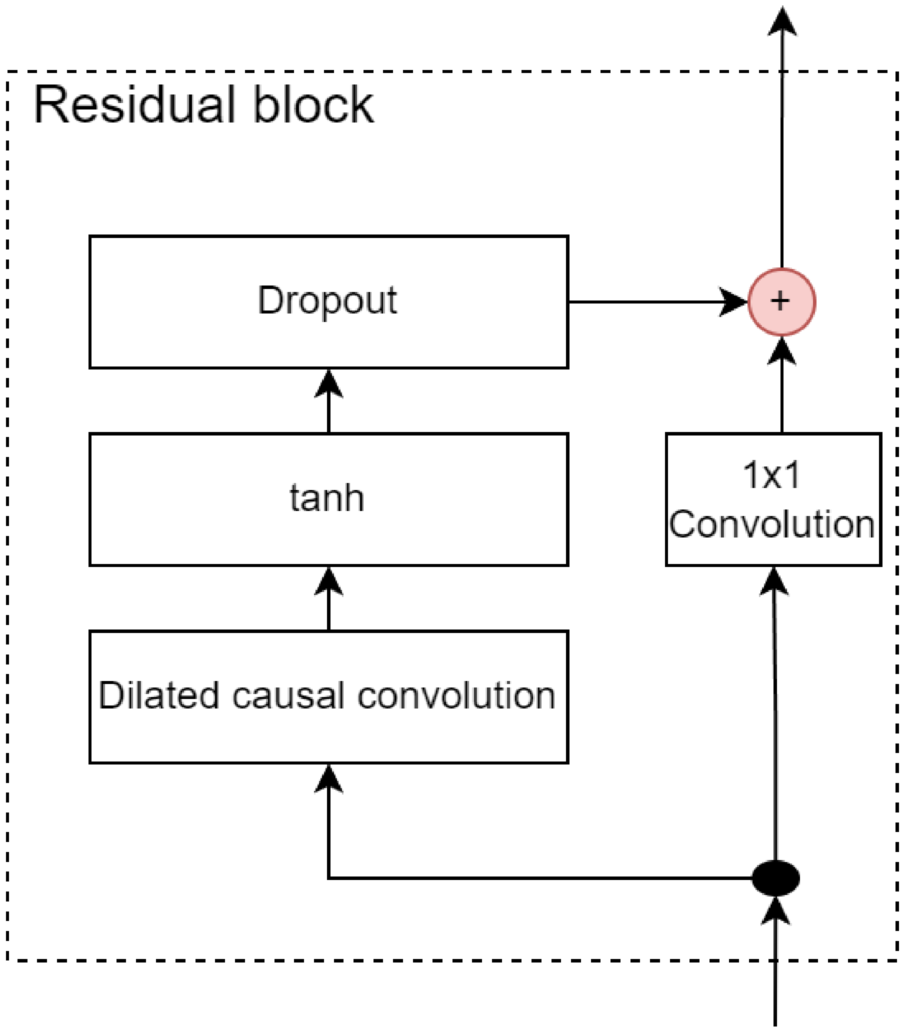

Figure 4.1: Residual block used by TCN

4.1.3 RNN, GRU and LSTM

The RNN model consists of RNN layers, where the layer combines the current input with the previous hidden state through learned weights, and an activation function, updating the hidden state for the next step. The layers use 24 units, adam optimizer with a learning rate of 0.003, tanh activation function and mean squared error loss function. Between each RNN layer is a dropout layer to regularize the result to reduce over-fitting in the model. The final layer is a dense layer which also uses tanh. The model also allows for the use of residuals, where the input to a layer is also added to the output for the layer. The model takes in a tensor and outputs a single value for the predicted power 24 hours later.

For hyperparameter optimization, the activation function, optimizer and units where all tested in an early version of the model and the ones which gave the best results then were chosen as the ones to continue to work with. The model was also tested for many different number of RNN layers with the conclusion being that the number of layers did not affect the performance directly except for beginning to over-fit when lots of layers were used. It affected how volatile the results were however which resulted in 3 layers being used as then the results were generally stable as well without over-fitting much.

The GRU and LSTM models are exactly the same as the RNN model except that they use GRU or LSTM layers instead of RNN layers.

4.1.4 24-hour Lag Model

A simple 24-hour lag model was also created to be used as a baseline. This model predicts that the energy consumption will be the same as the energy consumption 24 hours prior.

4.2 Model Performance

In this section, the performance of the different models will be presented. Comparisons were made between the number of residual blocks, the number of hours of input data, and which input features yielded the best performance, among others. The performance is measured by the root mean squared error (RMSE) when predicting the power consumption 24 hours ahead on the test set. However, because the test set varies depending on the number of input hours (due to differing lengths at the end of the dataset becoming unusable), the performance improvement over the baseline was added to better compare the different scenarios. For brevity, the input feature set consisting of Power, Temperature, and Observed busyness is referred to as "All".

Figures 4.2 and 4.3 below plot the predictions of the best performing TCN and RNN configurations.

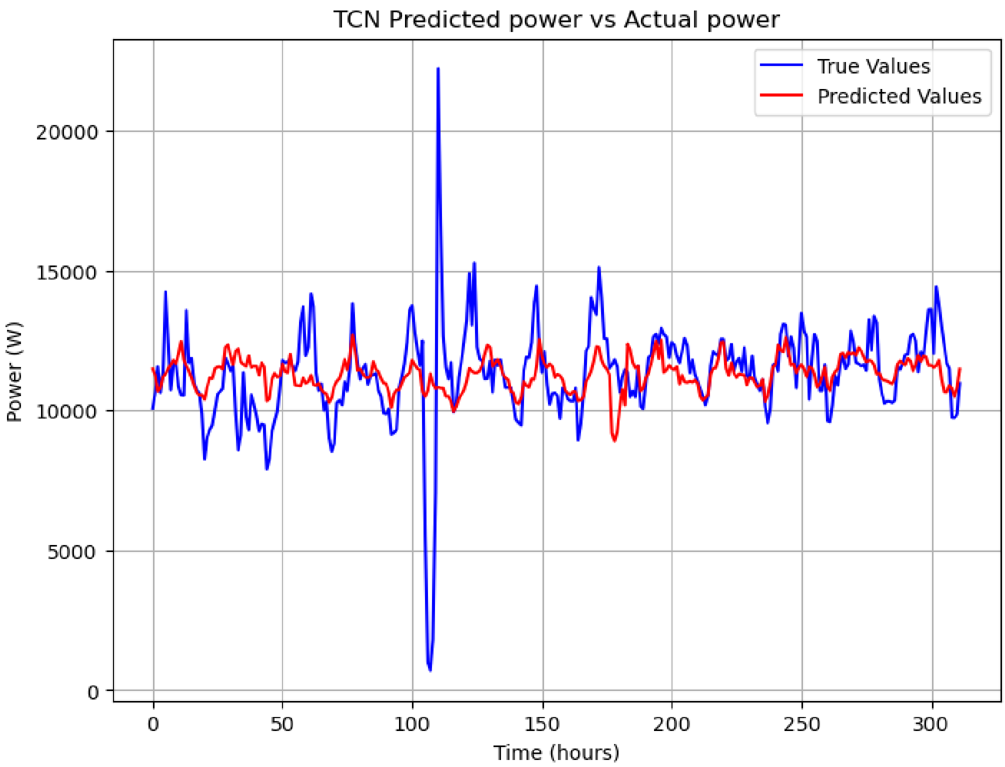

Figure 4.2: TCN predictions plot

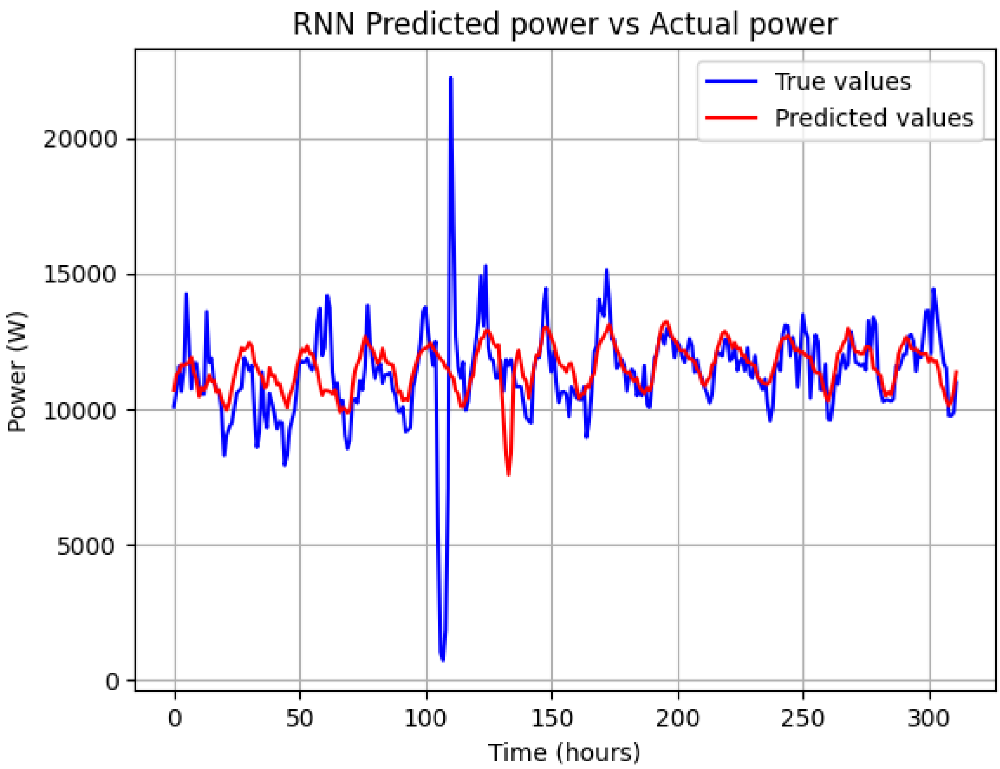

Figure 4.3: RNN predictions plot

Features	Blocks	Hours	24h RMSE	Improvement vs Baseline
All	3	48	1703.2	22.9%
All	1	48	1760.3	20.3%
All	5	48	1761.8	20.2%

Table 4.1: Performance of TCN for different numbers of residual blocks

4.2.1 Number of Residual Blocks

In table 4.1 the performance of the TCN model for different numbers of residual blocks is presented. The results are fairly similar regardless of the number of blocks, but it could be the case that three residual blocks performs better than more or fewer blocks.

4.2.2 Input Features

The performance of the models were compared for different input features and the results can be seen in table 4.2, 4.3, 4.4, and 4.5. below. The models consistently achieve very slightly better accuracy when using power, temperature, and observed busyness, than when using only power.

When comparing the observed busyness and the predicted busyness with the TCN model, it seems that the model using the observed busyness performs better.

Features	Blocks	Hours	24h RMSE	Improvement
Power	3	48	1749.0	20.8%
Power, Temp	3	48	1709.5	22.6%
Power, Temp, Observed busyness	3	48	1703.2	22.9%
Power, Temp, Predicted busyness	3	48	1774.5	21.3%

Table 4.2: Performance of TCN for different input features

Features	Layers	Hours	24h RMSE	Improvement
Power	3	48	1748.8	20.8%
Power, Temp	3	48	1748.0	20.8%
Power, Temp, Observed busyness	3	48	1702.8	22.9%

Table 4.3: Performance of RNN for different input features

Features	Layers	Hours	24h RMSE	Improvement
Power	3	48	1773.9	19.67%
Power, Temp	3	48	1713.4	22.4%
Power, Temp, Observed busyness	3	48	1686.9	23.6%

Table 4.4: Performance of LSTM for different input features

Features	Layers	Hours	24h RMSE	Improvement
Power	3	48	1782.1	19.2%
Power, Temp	3	48	1777.4	19.5%
Power, Temp, Observed busyness	3	48	1720.0	22.1%

Table 4.5: Performance of GRU for different input features

4.2.3 Number of Hours of Data

The models were tested with differing amounts of input data, the results of which are presented in table 4.6, 4.7, 4.8, and 4.9. For RNN, LSTM, and GRU, the amount of input data does not affect the performance.

However, for TCN, it is obvious that using 48 hours or more leads to significantly better performance than 24 hours. It does not seem like using more than 48 hours of data is superior, as the performance between 48, 72, and 168 hours is effectively the same. It should be noted that the model using 24 hours of input data has a 30.5% improvement over the 24-hour lag model on the validation set, while it performs much worse on the test set. This could be a sign of overfitting.

Features	Blocks	Hours	24h test	Improvement vs Baseline
All	3	24	2381.4	7.7%
All	3	48	1703.2	22.9%
All	3	72	1769.6	20.8%
All	3	168	2011.6	19.9%

Table 4.6: Performance of TCN for different number of hours of input data

Features	Layers	Hours	24h test	Improvement vs Baseline
All	3	24	1643.7	23.0%
All	3	48	1702.8	22.9%
All	3	72	1752.2	22.5%
All	3	168	1838.7	22.3%

Table 4.7: Performance of RNN for different number of hours of input data

Features	Layers	Hours	24h test	Improvement vs Baseline
All	3	24	1710.2	21.9%
All	3	48	1686.9	23.6%
All	3	72	1711.6	23.0%
All	3	168	1750.9	22.6%

Table 4.8: Performance of LSTM for different number of hours of input data

Features	Layers	Hours	24h test	Improvement vs Baseline
All	3	24	1707.6	22.0%
All	3	48	1720.0	22.1%
All	3	72	1709.5	23.1%
All	3	168	1776.3	23.3%

Table 4.9: Performance of GRU for different number of hours of input data

4.2.4 Residuals

The models were tested without residuals, and the results are shown in table 4.10, 4.12, 4.11, and 4.13. The performance with and without residuals is effectively identical. For RNN, the training time of the model using residuals is slightly shorter than the model with residuals. For TCN with residuals however, the training time was significantly shorter than without residuals. TCN with residuals reached its performance after roughly 130 epochs of training. TCN without residuals took around 20000 epochs to train.

Features	**Blocks**	**Residual**	**Hours**	**24h RMSE**	**Improvement**
All	3	Yes	48	1702.8	22.9%
All	3	No	48	1730.4	21.6%

Table 4.10: Performance of TCN with and without residuals

Features	**Layers**	**Residual**	**Hours**	**24h RMSE**	**Improvement**
All	3	Yes	48	1743.8	24.0%
All	3	No	48	1743.0	24.0%

Table 4.11: Performance of RNN with and without residuals

Features	**Layers**	**Residual**	**Hours**	**24h RMSE**	**Improvement**
All	3	Yes	48	1686.9	23.6%
All	3	No	48	1699.2	23.0%

Table 4.12: Performance of LSTM with and without residuals

Features	**Layers**	**Residual**	**Hours**	**24h RMSE**	**Improvement**
All	3	Yes	48	1720.0	22.1%
All	3	No	48	1706.3	22.7%

Table 4.13: Performance of GRU with and without residuals

5
Conclusion

This section contains a discussion of the results and also covers areas that can be improved upon in future work.

5.1 Discussion

By comparing the performance of the models for different input feature sets, it can be concluded that utilizing Google's busyness data slightly improves short term load forecasting accuracy. While it was expected that the predicted busyness might not be very useful due to its low accuracy, an attempt was made to mitigate this by using the observed busyness instead, and by doing so arriving at an upper bound of how useful the busyness metric could be, if its predictions were more accurate. The result shows that both the predicted and the observed busyness indeed lead to performance improvements, but the performance gains of using the predicted busyness seem to be much smaller.

Another conclusion one can draw from the results is that the outside temperature slightly improves short term load forecasting accuracy. This is to be expected, as the outside temperature affects the efficiency of the compressors, and prior studies have shown that the outside temperature correlates with the power consumption [8].

For TCN, using residuals significantly reduces the required training time, from 20000 epochs without residuals to 130 epochs with residuals. In fairness, the training time without residuals could likely be reduced significantly with the use of variable learning rates, but it would likely still take longer than when using residuals. Unlike the deep networks for image recognition that saw performance benefits from using residuals, residuals did not improve the performance of the models that were tested. Still, even for a small TCN, residuals seem to be a beneficial tool for reducing training time for short term load forecasting models. RNN, LSTM, and GRU on the other hand do not benefit as much or at all from residuals. This could reasonably be because they already incorporate memory mechanism. LSTMs and GRUs especially, with their advanced gating mechanisms, are particularly effective at maintaining and regulating information over long sequences and therefore need residuals even less than RNN.

To summarize, in this thesis we have used machine learning models to predict the power consumption of a supermarket refrigeration system 24 hours in advance, which means that it would be possible for supermarkets to use similar models in order to bid on the energy flexibility market, and thus increase the energy flexibility of Sweden. The results show that temperature is a useful parameter for enhancing prediction accuracy, but also that the number of people in the store, or the number of cabinet interactions, could have an even bigger potential to enhance the models.

5.2 Future work

In this thesis, it has been shown that using Google's popularity metric can improve the short term load forecasting performance. However, the popularity is only a proxy for the number of people in the store, which itself is a proxy for the number of cabinet interactions. Therefore, since all Swedish stores already track how many people are in the store at all times, a reasonable next step would be to use the number of people in the store instead of the busyness. Furthermore, investigating how the number of cabinet interactions affects the power consumption, rather than approximating the number of cabinet interactions through the number of people in the store, would also be useful.

After seeing the temperature data, it was believed that the inside temperature would not have a large effect on the performance of the models, due to the relatively consistent temperature. However, since the data for the inside temperature was corrupted it was not possible to test this hypothesis properly. Therefore, a future area to explore would be if the inside temperature impacts the short term load forecasting accuracy or not. On one hand, the inside temperature should have a large impact on the power consumption of the refrigeration system, but on the other, it seems to not fluctuate very much, in which case it would not impact the forecasting accuracy.

Due to the structure of the convolution in TCN and other convolutional models, with each filter always applying to the same indices in the input data, theoretically it could be beneficial if the indices always corresponded to the same hour of the day. This would hypothetically lead to better consistency for the data the filters are operating on, and therefore better prediction accuracy. While some early testing was done with this method, it reduced the size of the training set too much to be able to draw any reasonable conclusions. Therefore, investigating this on a larger dataset could be of value.

The models in this thesis were trained on 3 months of data, which means that the dataset captures weekly and monthly trends. However, it is possible that the power consumption of supermarket refrigeration systems follows yearly trends, which can not be captured in such a small dataset. While none of the models that were used in the thesis can capture such long seasonality, a larger dataset could possibly allow for the use of models which can capture longer trends.

Bibliography

[1] Sveriges miljömål, "Andel energi från förnybara energikällor," [Online]. Available: `https://www.sverigesmiljomal.se/miljomalen/generationsmalet/fornybar-energi/` (visited on 01/2024).

[2] Svenska kraftnät, "Lagring av el - omvärldsanalys," Nov. 2022. [Online]. Available: `https://www.svk.se/siteassets/om-oss/rapporter/2022/rapport-ru-energilager.pdf` (visited on 01/2024).

[3] J. Arias, "Energy usage in supermarkets: Modelling and field measurements," Ph.D. dissertation, 2005.

[4] T. Månsson, A. Sasic Kalagasidis, and Y. Ostermeyer, "Hygro-thermal model for estimation of demand response flexibility of closed refrigerated display cabinets," *Applied Energy*, vol. 284, p. 116 381, 2021. DOI: `10.1016/j.apenergy.2020.116381`.

[5] H. Pihl, "Swedish fcr prices – an analysis of the data," 2019. [Online]. Available: `https://www.ri.se/sites/default/files/2021-02/Swedish-FCR-prices_1.pdf` (visited on 01/2024).

[6] Svenska kraftnät, "Resultat från pilotstudie för resurser med begränsad energireserv," 2023. [Online]. Available: `https://www.svk.se/aktorsportalen/bidra-med-reserver/bli-leverantor-av-reserver/bidra-med-fcr-afrr-eller-mfrr/delta-pa-fcr--marknaderna-med-resurser-med-begransad-energireserv--ler/resultat-fran-pilotstudie-for-resurser-med-begransad-energireserv/` (visited on 01/2024).

[7] Livsmedelsverket, "Säker mat i din butik," Mar. 2018. [Online]. Available: `https://www.livsmedelsverket.se/globalassets/foretag-regler-kontroll/branschriktlinjer/butik---saker-mat-i-din-butik.pdf` (visited on 01/2024).

[8] L. B. Rasmussen, P. Bacher, H. Madsen, H. A. Nielsen, C. Heerup, and T. Green, "Load forecasting of supermarket refrigeration," *Applied Energy*, vol. 163, pp. 32–40, 2016. DOI: `10.1016/j.apenergy.2015.10.046`.

[9] T. Månsson, A. Rukundo, M. Almgren, P. Tsigas, C. Marx, and Y. Ostermeyer, "Analysis of door openings of refrigerated display cabinets in an operational supermarket," *Journal of Building Engineering*, vol. 26, p. 100 899, 2019.

[10] R. T. Faramarzi, B. A. Coburn, and R. Sarhadian, "Performance and energy impact of installing glass doors on an open vertical deli/dairy display case," *Ashrae Transactions*, vol. 108, p. 673, 2002.

[11] A. Gasparin, S. Lukovic, and C. Alippi, “Deep learning for time series forecasting: The electric load case,” *CAAI Transactions on Intelligence Technology*, vol. 7, no. 1, pp. 1–25, 2022.

[12] Open-Meteo. [Online]. Available: `https://open-meteo.com/en/docs/historical-weather-api/` (visited on 01/2024).

[13] C. Lea, R. Vidal, A. Reiter, and G. D. Hager, “Temporal convolutional networks: A unified approach to action segmentation,” in *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14*, Springer, 2016, pp. 47–54.

[14] S. Bai, J. Z. Kolter, and V. Koltun, “An empirical evaluation of generic convolutional and recurrent networks for sequence modeling,” *arXiv preprint arXiv:1803.01271*, 2018.

[15] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[16] R. M. Schmidt, “Recurrent neural networks (rnns): A gentle introduction and overview,” 2019. arXiv: `1912.05911 [cs.LG]`.