Neural Network-Based Modelling of the Thermal Dynamics for SBB Trains and their HVAC Systems

Master Thesis

Author(s):
Parravicini, Diego

Publication date:
2024-06-30

Permanent link:
https://doi.org/10.3929/ethz-b-000704869

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Automatic Control Laboratory

Master's Thesis

Neural Network-Based Modelling of the Thermal Dynamics for SBB Trains and their HVAC Systems

Diego Parravicini
June 30, 2024

Advisors
Prof. Dr. John Lygeros
Dr. Ahmed Aboudonia
Dr. Raffaele Soloperto

Abstract

Over the course of the last year, the Swiss Federal Railways displayed an annual energy consumption of 2300 Gigawatt-hours, with this number, as it stands, destined to grow since the increasing demand for public transportation in Switzerland.
Between 15 to 20% of the power demand of the SBB fleet trains is due to Heating, Ventilation and Air Conditioning systems which is responsible for the well-being and comfort of the passengers and has to comply with strict norms.

In order to achieve monetary and emission savings, one of the option is a more precise and smart deployment of the HVAC Systems: the implementation of a Model Predictive Control (MPC) techniques could lead to these goals.
This study aims to investigate the feasibility of modelling the thermal dynamics of a wagon of the Regio-Dosto train through Neural Networks with the future possibility to use it within a MPC framework.

Data from 8 trains during the Winter period and 5 trains during the Summer period are used in a simulation environment to train and test different Neural Networks architecture, taking into account the non-linearities affecting the thermal models of the wagons.
In particular, a novel type of Physics informed Neural Networks (PiNNs), named Physically Consistent Neural Networks (PCNN), are used to simulate the complexity of this model using not only the amount of past data retrieved by SBB over the course of past months, but also the basic physics equations behind these processes. This novel architecture can lead to a convex input-output relationship that can be utilized in a future optimization step.

Results show the effectiveness of this model architecture: different architecture typologies of the same kind can be exploited in order to make progress towards a potential implementation of MPC in the SBB fleets.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Energy Consumption of the Swiss Federal Railways (SBB CFF FFS)

According to the official Data from the Swiss Federal Railways [24] in 2023 an average of more than 7800 trains operated each day: this represents a +4.6% increase compared to 2022 and surpasses pre-COVID-19 levels, which saw a decline over the three-year period from 2020 to 2022.
Out of a population of nearly 9 million people, SBB reports 1.32 million passengers per day, the exact same number of 2019. Moreover, as shown in Figure 2.1 with a comparison between the year 2022 and 2019, Switzerland is the country with the highest passenger transport demand in Europe, reaching almost 2250 kilometers traveled by rail per inhabitant in 2022.

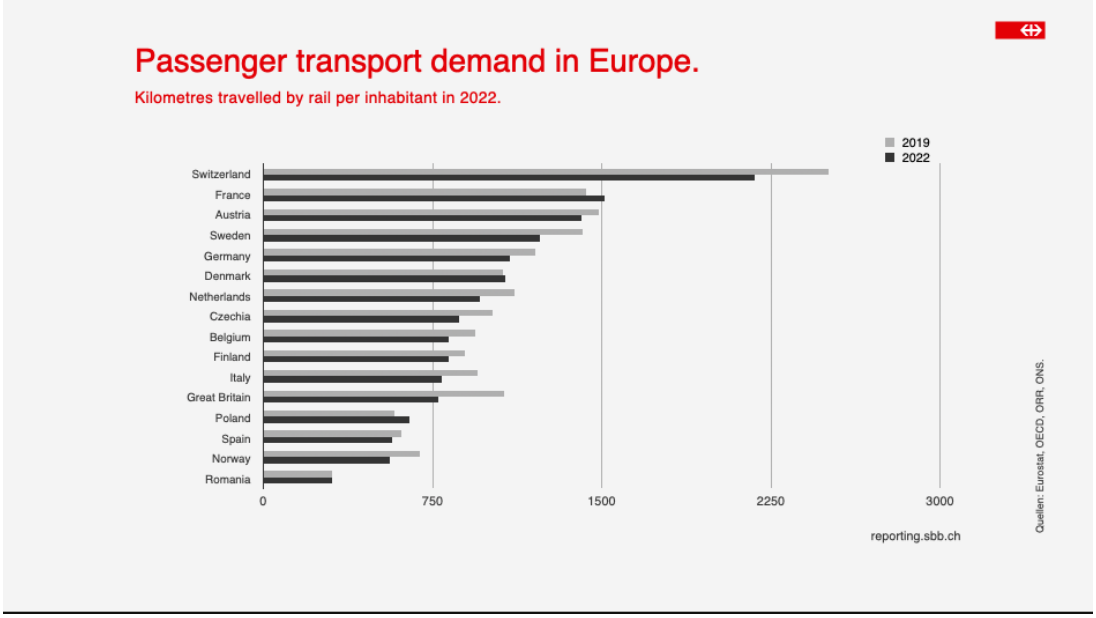


Figure 1.1: Passenger Transport Demand in Europe, comparison between 2019 and 2022. Sources are Eurostat, OECD, ORR, ONS, as indicated on the right side of the picture

In 2023, the electricity and fuel consumption of the rail network in Switzerland amounted to over 2300 gigawatt-hours [24], equivalent to the energy use of 460,000 households, assuming an average consumption of 5,000 kWh as reported by the Swiss Federal Office of Energy (SFOE) in 2021 [11].
Altogether, this consumption results in 62,000 tonnes of greenhouse gas emissions. Over 70% of the 2,300 gigawatt-hours is utilized solely for traction current (Figure 2.2).

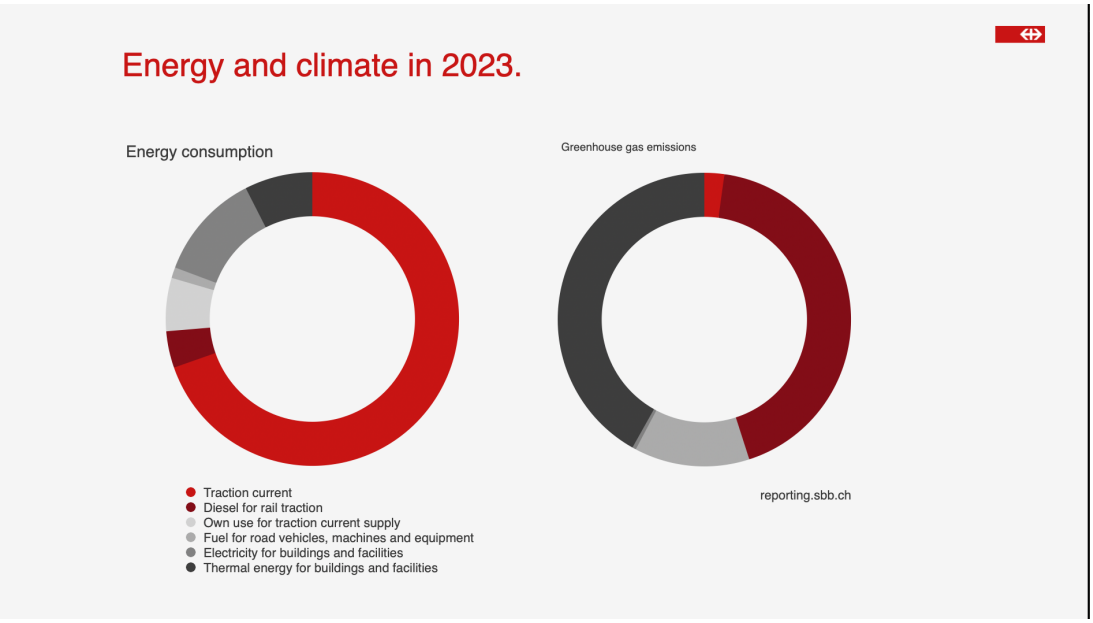


Figure 1.2: Energy Consumption and Greenhouse Emissions of the railway network in 2023

Energy Efficiency and Sustainable Energy are keywords for SBB's future strategy [23]. In the near future, SBB's primary aims include:

- saving up to 20% of its forecast annual energy consumption for 2025, or a total of 600 gigawatt hours (GWh);
- running trains on electricity generated entirely from renewable energy sources starting in 2025;
- deploying smart grids for an optimal Load Management;
- improving its energy efficiency by around 30% by 2030 compared with 2010;
- promoting new sources of renewable energy and so reducing greenhouse gases.

1.2 Energy Consumption of a SBB train and the HVAC System

The energy consumed by trains is mainly distributed as follows [15]:

- traction;
- auxiliary traction operations (compressors, fans);
- HVAC (heating, ventilation & air conditioning);
- small consumers (lighting, 4G repeater, etc.);

Thanks to a measurement project in collaboration with the University of Basel, relatively accurate energy values are available from the ICN (RABe500), which can serve as a good example. Traction consumes more than 70% of the total energy deployed. The HVAC System, instead, amounts to 15-20%.
The latter encompasses all the equipment necessary to regulate temperature, humidity, airflow and air quality in commercial and industrial buildings, as well as in various modes of transport, with the primary aim of ensuring thermal comfort for occupants.

In a "Synthesis Report" of 2021 issued by the Federal Office of Transport to the Lucerne University of Applied Sciences and Arts [13], experts analyzed more than one hundred papers pertinent to the air conditioning energy saving for public transportation. These recent studies indicate that approximately 20-40% of the energy is used for heating, ventilation, and cooling (HVAC) in passenger compartments of vehicles (trains, buses, trams), while 60-80% of the energy is required for traction.
They also drafted a list of the eight measures most commonly cited for their high potential, which are the following:

- Indoor temperature adjustment
- CO_2-controlled ventilation
- Sleep mode
- Heat pump use
- Exhaust air heat recovery
- Model-predictive control
- Better windows (lower heat transfer coefficient)
- Insulation

All of these measures are believed to have a savings potential of 5-30%. The significant variation in estimates is likely due to factors such as vehicle type, usage patterns, and outdoor climate, all of which greatly impact the achievable potential.

1.3 Artificial Neural Networks (ANN)

Artificial Neural Networks are computational models made by an interconnected group of nodes, which mimics the biological behavior of the working units of the brain, the neurons, and their connections, the synapses. Their use, with the constant growth of interest for Artificial Intelligence and Machine Learning in the last decades, spans from Natural Language Process to Autonomous Systems.
Among the most famous Artificial Neural Network Architecture, it is possible to find:

- FeedForward Neural Networks (FFN): the connections between Neurons don't form cycles and the information flows only in one direction
- Convolutional Neural Networks (CNN): a type of Feed-Forward Neural Network which makes use of filter optimization. They are mostly used for image and video recognition, natural language processing and recommended systems
- Recurrent Neural Networks (RNN): the connections between neurons happen in a cycle way

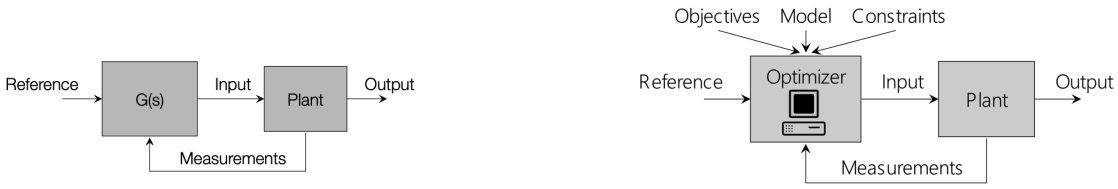
Because of this inherently complex internal structure, Neural Networks can be seen as Black-Box models, systems which can be identified mainly by their inputs and outputs, without a clear knowledge of the patterns that connect them. A deeper dissertation between Black-Box Models and their difference with Physics-Based Models is held in the 3.6.1 section.

1.4 Model Predictive Control (MPC)

Model Predictive Control (MPC) is an advanced control strategy used in various industrial and engineering applications, such as Robotics, Automotive and Energy Systems.
MPC uses a dynamic model of the system, which can be linear or nonlinear depending on its complexity, to predict its future behavior over a specified prediction horizon and compute the optimal sequence of control inputs to be applied.
At each control step, an optimization problem is solved to determine the control inputs that will minimize a cost function, also taking into account constraints on inputs, states and outputs. These constraints can include physical limitation, operational limits, safety or, like in the case of the HVAC Systems, comfort requirements.
Compared to a more classical control method, like the proportional-integral-derivative controller (PID controller), where the control inputs are automatically adjusted based on the data and not based on a model and constraints are not considered, MPC proves to be more demanding from a computational standpoint, but also better in terms of cost effectiveness and energy efficiency [3].

1.5 MPC in Building Climate Control: a similar case

The theory of comfort for public transport originated from building comfort theory [13].
It is estimated that one third of the CO_2 emissions comes from buildings [18], for which HVAC results as one of the major contributors. MPC turned out to be a valid data-driven option in the context of heating and cooling energy saving processes [12], but developing effective MPC-based control strategies for HVAC systems is nontrivial since buildings dynamics are nonlinear and influenced by various uncertainties [7].
To this end, the Automatic Control Laboratory at ETH Zurich (*IfA*), explored the potential implementation of MPC for HVAC Systems in buildings, specifically the NEST, a modular



(a) General Feedback Control Scheme

(b) General MPC-based Control Scheme

Figure 1.3: Comparison between a classical Control Loop (a) and an MPC based Control Loop (b). The controller *G(s)* is replaced by an optimizer which utilizes a Model, Objectives and Constraints [19]

research and innovation building at EMPA [2]. Compared to a traditional hysteresis controller, this approach achieved a 25% reduction in cooling energy consumption and a 72% decrease in the integral of comfort constraint violations over a six-day experiment conducted in a simulation environment [4].
Another study [6] showed how three different machine learning based predictive control methods for building climate control can be efficient and outperform the baseline controller.

Figure 1.4: NEST Building, Zooey Braun, Stuttgart.

1.6 Previous Works

In the past semesters, a bachelor thesis, a master thesis and some related detailed studies about HVAC and energy saving were carried out by students and colleagues of the Automatic Control Laboratory at ETH Zurich.
In his Bachelor Thesis [25] from the Spring Semester 2022, M. Schnetzer proposed a model predictive controller starting from a state-space model of the original physics-based model theorized by SBB in a Load-Management Report of 2019 [14]. The train in question was the SBB Regio-Dosto Train. This Model Predictive controller was then evaluated against the controller already implented on the trains: a hierarchical control system architecture with Rule Base Controller (RBC) at the top level and a proportional- integral-derivative (PID) controller at the bottom

level. The comparison setup involved a case study utilizing a combined dataset spanning three days and three different routes, incorporating realistic operating conditions.
The case study pointed out a potential energy saving for the HVAC systems equal to 11% while maintaining a similar performance in terms of constraint violations. Some inconsistencies in the results, though, highlight certain limitations of the developed MPC.
N. Wertli, in his Master Thesis [27] from the Winter Semester 2023, investigated the advantages of using an ARX-type (Autoregressive model with exogenous input) based model for prediction controllers with the same scope of M. Schnetzer, using the same model of train as he did, the Regio-Dosto. Two different MPC options were analyzed: a "Full MPC" with direct control of the individual heat flows and a high-level MPC combined with the PID as a low level controller. Assuming a perfect forecast of disturbances in the model, these two Model Predictive controllers were tested against the Rule Based Controller. The "Full MPC" option resulted in a 5-10% energy saving and was able to perform better than the baseline controller also in the scope of constraint violations: this was mainly thanks to the ability of tracking a temperature trajectory nearer to the energy-efficient tolerance limit and not in the prediction part itself. The high-level MPC, instead, resulted in a 5.2% energy saving but fails to meet the temperature constraints as consistently as the baseline controller.
Building upon the work of N. Wertli, Dr. Aboudonia continued to investigate the potentiality of the linear ARX-Model [26]. The current implementation uses measurements from the last 60 minutes (named *lags*), while the prediction horizon is 30 minutes.
In this case the energy savings in the winter period amount to 10.55-11.25% and in the summer period to 7.4-7.8% compared to the usual baseline controller.
Based on the estimates during the ETH-SBB Workshop in April 2024 [26], it is expected a total yearly savings of approximately 150000 to 250000 CHF or 1-2 GWh. This would scale up with more trains using MPC.

1.7 Goal of the Thesis

As specified in the previous section, SBB highlights the commitment towards a more sustainable way of traveling and resources management.
The power consumption resulting from the Heating, Ventilation and Air Conditioning of trains fleet has a great impact on the overall emissions and reducing this contribution could result in major benefits for the company itself and the environment.
The goal of this Master Thesis is, building on the work of the students and colleague from the IfA Laboratory at ETH„ is exploring for the first time the chance to model the internal thermal dynamics usign Neural Networks in a simulation environment.
Starting from here, the natural continuation could be implementing a Model Predictive Controller based on the aforementioned model and eventually run the controllers in a real environment.

Chapter 2

Considered System

2.1 SBB Regio-Dosto

Among the SBB's Passenger Traffic fleet, the RABe 511, or the Regio-Dosto Train, is a double-decker train used both on the S-Bahn lines of the Zurich S-Bahn and on the IR lines in the Zurich area and in Western Switzerland [21].

The manufacturer is the Swiss Company Stadler Rail, and this model is part of the Stadler KISS family of trains which is the acronym in German for "Komfortabler, Innovativer, Spurtstarker S-Bahn-Zug", meaning "comfortable, innovative, sprint-capable suburban train" [28]. It is a well-known suburban train also outside of Switzerland: as a matter of fact, it is employed in countries like Germany, Austria, Russia, Luxembourg and Georgia.

Launched for the first time in 2010, the current number of these trains is 56 and they have 535 seats per unit if they come with the 6 wagons composition (150 meters long) or 337 seats if in the 4 wagons composition (100,36 meters long). They can reach a traction force of 400 KN, a continuous output at wheel of 4000 kW and a maximum speed of 160 km/h.

Figure 2.1: Exterior Views of the RABe 511 [22]

Figure 2.2: Interior view (lower deck, Second Class) of the RABe 511 [22]

2.2 Air Conditioning and HVAC System

As stated in the SBB Load Management report [14] and in the Air Conditioning document from Stadler and SBB [16], the wagons of the RABe 511 are symmetrical, and from now on only half of it will be considered for the Thesis work and assumed that the two halves behave similarly. Inside the half wagon, three distinct parts can be identified, which correspond to three different thermal zones:

- Upper deck
- Middle deck
- Lower deck

Heating does not only come from the HVAC units mounted in the passenger carriage. In the Figure 3.3, apart form the three decks specified above, it is possible to observe:

- the floor heating zones (yellow)
- the wall heating zones (orange)
- the heater blowers (blue) and the foot-board heating zones (green), both of them close to the entrance to provide the necessary heat flow to maintain an adequate temperature during the door opening times in Winter [16].

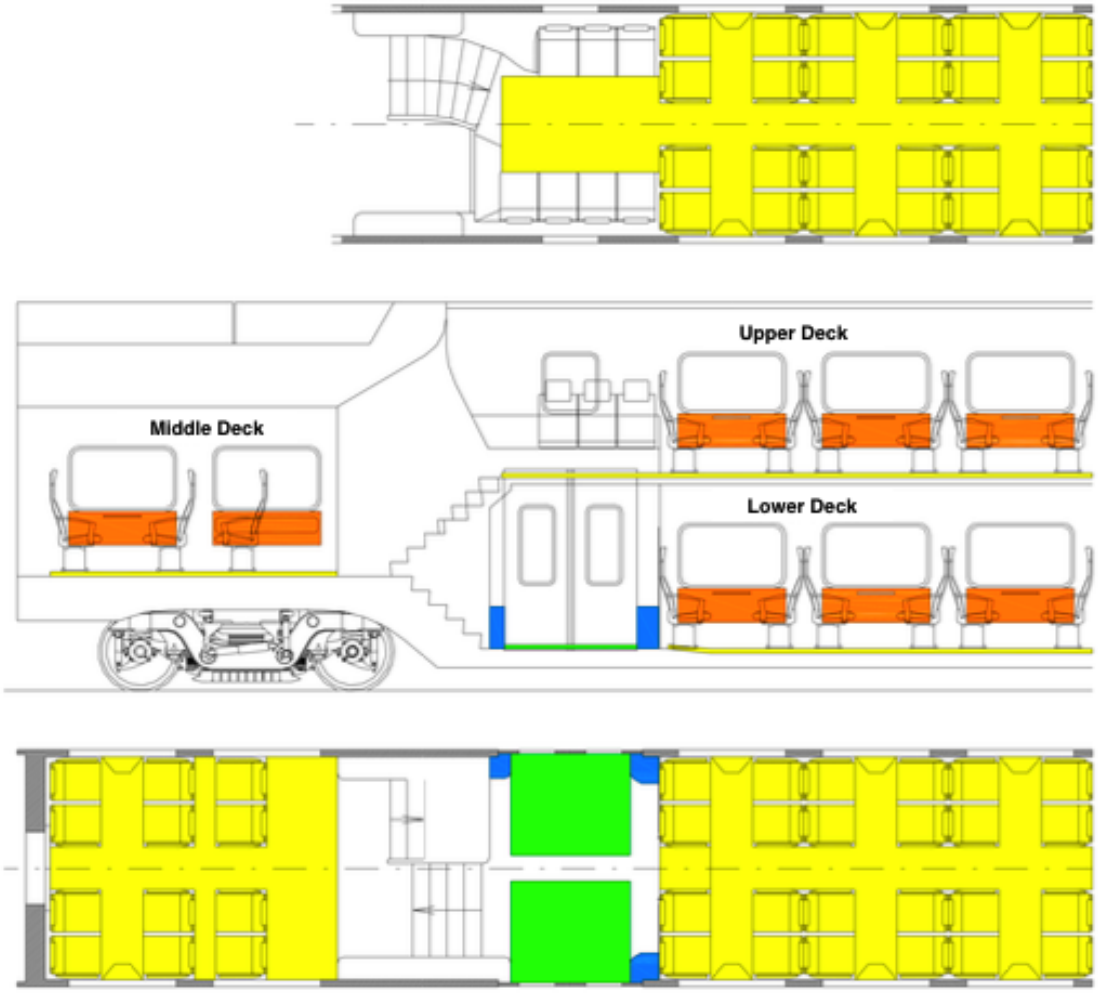


Figure 2.3: Longitudinal section of a wagon of the RABe 511 [22]

2.3 HVAC Unit and its functioning

Each half of the wagon is equipped with one HVAC unit, mounted on the ceiling of the carriage. The upper, middle, and lower decks are each controlled by separate HVAC controllers or modules within the same unit, ensuring that their ventilation, heating, and air conditioning, although partially influenced by one another, are managed individually. These three HVAC controllers also handle the actuation of the respective floors and walls.
Each of the three modules is equipped with an electrical heating element (used for the heating scenario), a refrigeration circuit (for the cooling scenario), an air supply fan and flaps for outside air or recirculated air.

The air flow, which is made of a combination of re-circulating air and outside air, is represented in blue in Figure 3.4. In red the heat flows from wall and floor heating. Based on the number of the passengers, the right amount of fresh air is mixed with the recirculated one to ensure the right ventilation comfort for the occupants.
In contrast to recirculated air, the outside air must first be heated or cooled to room temperature in winter or summer, while the recirculated air is already available at this temperature. In the RV-Dosto, the outside air volume is controlled depending on the occupancy level and the outside air temperature

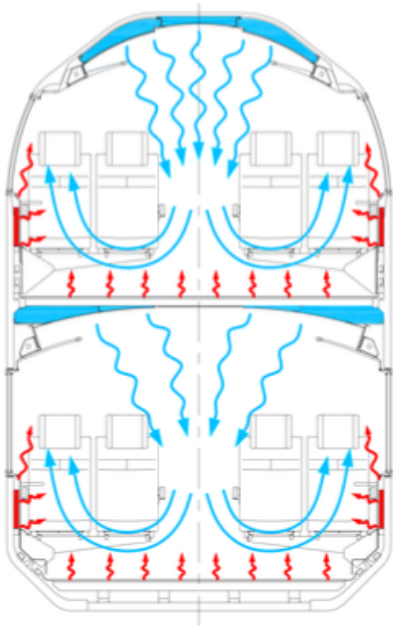


Figure 2.4: Front section of the wagon that shows air and heating flows [16]

The HVAC unit manufacturer is Faiveley Transport, an international company and supplier of equipment for the railway industry [8]. More detailed information of the HVAC unit and its sketch can be found in the Appendix section.

2.4 Standards and Regulations for Passenger Comfort

Among the various regulations that must be met, those of greatest relevance to this thesis are EN 14750-1 and EN 14750-2. These European Standards, introduced in 2006, are applicable to suburban and/or regional vehicles and also metro and tramway vehicles equipped with cooling and/or heating/ventilation systems. They are mostly related to Fresh air volume supply, Air velocities and Overall sound level [10].
Some additional requirements to these European Standards have be to applied in compliance with SBB and Switzerland norms [20].
In particular:

- The deviation of the set-point room air temperature from the average room air temperature at 1.1 meters above the floor is limited to a maximum of 1K
- The maximum deviation of room air temperatures at 1.1m above floor is restricted to 2K

- The new set-point temperature must not deviate by more or less than 2K from the set-point curve specified by SBB
- It should be possible to reach the new set-point temperature within 15 min/K

2.5 Modeling of the Thermal Dynamics of the Regio-Dosto

A Predictive Thermal Load management [14] was conducted by Jan Gasser from SBB with the main goal of investigating the peak shifting potential of the HVAC systems.
SBB already operates the so called "load management": controlling and manipulating of electricity consumption to ensure supply-demand balance in a power grid, especially when this grid presents high and frequent fluctuations.
The report, eventually, showed that with appropriate control, all air conditioning systems in the SBB vehicles could make a significant contribution to smoothing the half-hourly cycle variation without compromising the temperature tolerance in the passenger compartments.
This study is of particular interest for the scope of this Thesis, because of the modeling of the HVAC systems, the heat flow dynamics and a MATLAB Simulink model that were developed. Most of the assumptions on the model, like the one of the halves of the wagon behaving similarly, is inspired by this study report.

2.5.1 Basic Equations of the Model

The thermal dynamics of the wagon are modeled as an RC model made of different heat accumulators connected to each others and that exchange heat in different ways. For each of them a differential equation describes the temperatures gradients based on the the exchange of heat flows.

$$\frac{\partial T_i}{\partial t} = \frac{1}{m_i \cdot c_i} \cdot \sum_{j=1}^{N} \dot{Q}_j \tag{2.1}$$

T_i is the temperature of the specific heat accumulator, m_i is its mass in kg and c_i is the heat capacity in $J/(kg \cdot K)$. $\dot{Q}_j$ is the heat inflow or outflow related to the heat accumulator, in the first case it is positive and in the second case negative. It is measured in J/s.
To calculate the heat exchange between two accumulators the following equations is assumed to hold:

$$\dot{Q}_{1,2} = k_{1,2} \cdot A_1, 2 \cdot (T_1 - T_2) \tag{2.2}$$

In this equation $A_1, 2$ represent the area between the two accumulators, and k is a heat transfer coefficient which will be discussed in more details in section 3.5.3.

2.5.2 RC diagram of the model

The passenger carriage is modeled using a total of nine different accumulators while the ambient is considered as an infinitely large reservoir. In addition to the upper deck, middle deck and lower deck of the half wagon, also the chassis and the inventory (seat and so on) of each of these decks are considered.
As shown in figure 3.5, between the nine main thermal compartments the heat flows are transmitted through conduction, convection, air circulation and radiance. The yellow arrows show that radiance transmission is considered only between the ambient and the inside, but not between inside accumulators.

Arrow's color	Heat Transmission Mode
Purple	Conduction
Yellow	Irradiation
Blue	Airflow
Green	Convection

Table 2.1: Heat Transmission Modes.

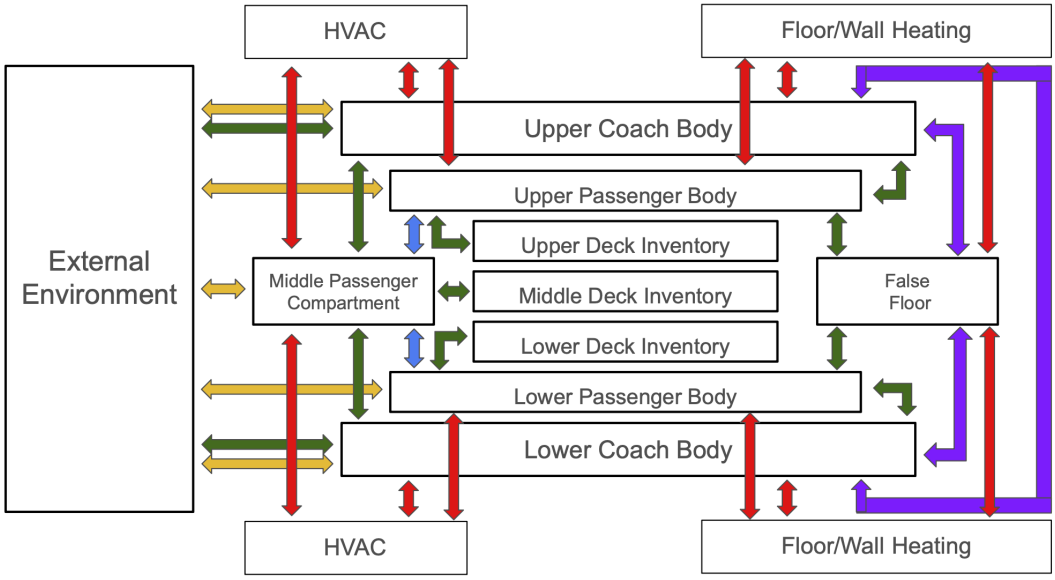


Figure 2.5: RC Diagram of the Model with 9 accumulators and the different modes of heat transmission. While the purple, yellow, green and blue arrows represent the Heat Transmission Modes, red arrows show the areas influenced by the HVAC and floor/wall heating systems.

2.5.3 k-values estimation

During the report assessment, SBB requested Klimatechnik to estimate the k-values, a parameter that describes the thermal resistance between the interior and the external environment and takes into account all type of of heating transfer modes. This value was estimated as a function of the speed of the train when the HVAC system is in heating mode. The report also displays the k-values for different trains of the SBB fleet.

speed (km/h)	k-value $(W/(m^2K))$
10	1.66
60	2.05
120	2.29

Table 2.2: k-values depending on the speed of the Regio-Dosto.

The table 3.2 shows a strict correlation between the speed and the heat exchange between the inside and the ambient.

2.6 Modeling of the HVAC System

The HVAC system can have three possible states:

- slumber mode
- automatic mode
- off mode

In this thesis, the focus will be on the automatic mode, where the HVAC system operates regularly, aiming to reach the temperature reference set by the controller, as will be detailed later. However, it is also worth discussing the functioning of the slumber mode.
This mode was primarily intended to save energy, a middle way between the automatic and off mode. A temperature reference is not directly specified, only the internal temperature and the ambient temperature play a role in this mode. Its functioning is better explained in detail in the following table.

Condition	Operation
if $T_{room} < 10.5°C$	Heating On with 100% of circulation air
if $T_{room} < 24°C$ and if $T_{room} < T_{amb}$	Circulation On
if $T_{room} > 32.5°C$ and $T_{room} > T_{amb}$	Cooling on with 100% of outside air
if $T_{room} > 32.5°C$ and $T_{room} < T_{amb}$	Cooling on with 100% of circulation air
else	Off Mode

Table 2.3: Slumber mode cases. Notice how outside air is used if it is cooler than the air in the passenger compartment.

2.6.1 Control Mode

The regulation of the room temperature inside the Regio-Dosto train is operated by a hierarchical control scheme: an high-level rule-based controller is used to decide the temperature setpoint and a lower-level PID controller is used for setpoint tracking.

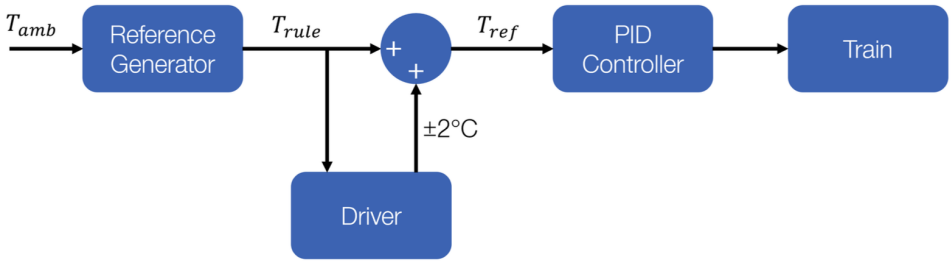


Figure 2.6: Current Hierarchical Control Scheme made of a Rule Based Controller and a PID Controller [27].

In figure 3.6 is also present the train driver block: the conductor, in fact, can manually increase or decrease the reference temperature by 2 ∘C through discrete steps of 0.5 ∘C.
The temperature reference, excluding the external intervention of the train conductor, solely depends on the ambient temperature. The relationship between the ambient temperature and the reference temperature is shown in Figure 3.7.

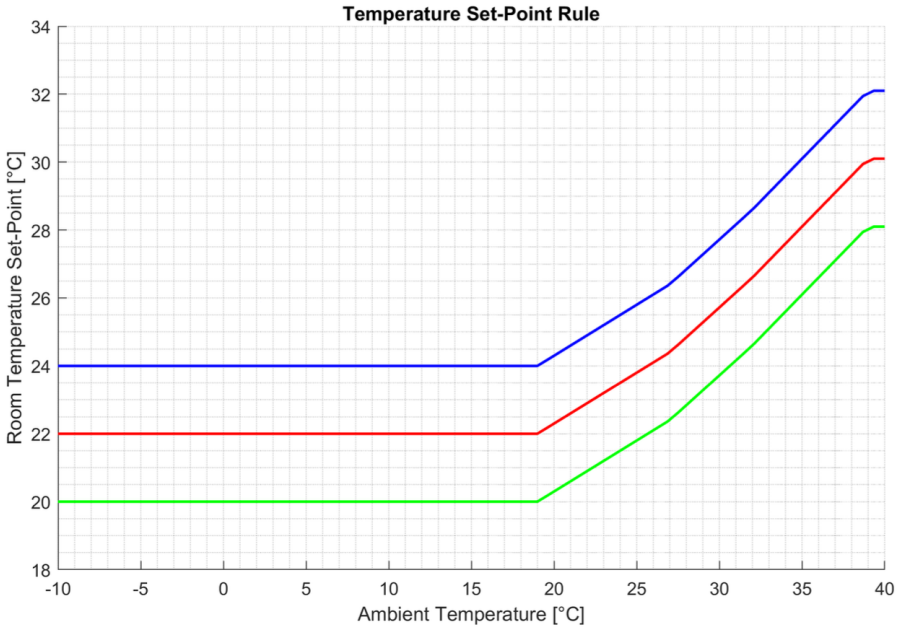


Figure 2.7: Relationship between the reference temperature and the ambient temperature [27]. The red line represent the actual rule, the blue and green lines represent the limit within which the driver can impose manual modifications.

The piecewise linear function in red is defined as follows:

- if $T\_amb < 19°\text{C}$ then $T\_ref = 22°\text{C}$
- else if $19 \leq T\_amb < 24°\text{C}$ then $T\_ref = \frac{24.4-22}{27-19} \cdot (T\_amb - 19) + 22$
- else if $27 \leq T\_amb < 32°\text{C}$ then $T\_ref = \frac{26.6-24.4}{32-27} \cdot (T\_amb - 27) + 24.4$
- else if $32 \leq T\_amb < 39°\text{C}$ then $T\_ref = \frac{30.1-26.6}{39-32} \cdot (T\_amb - 32) + 26.6$
- else $T\_ref = 30.1°\text{C}$

The range in which the temperature reference lies is between 22 °C and 30.1 °C and at the same time the Regulations in section 3.4 have to be respected.
Additional information about the PID Controller and its functioning are not disclosed by the manufacturer Faiveley.

2.6.2 Modeling of the floor/wall heating

Information about the operating modes of the floor/wall heating from the manufacturer were not available, therefore, by analyzing the temperatures, the heating capacity was reconstructed a a function of the ambient temperature.

$$\dot{Q}_{Floor/Wall} = -\frac{2}{5} \cdot T_{amb} + 3.5\ kW \tag{2.3}$$

With $dotQ_{Floor/Wall}$ representing the heat flow in the half wagon and T_{amb} the outside temperature.
For example, in the case of heating mode, the relation between the ambient temperature and the percentage of the heating power is the following one:

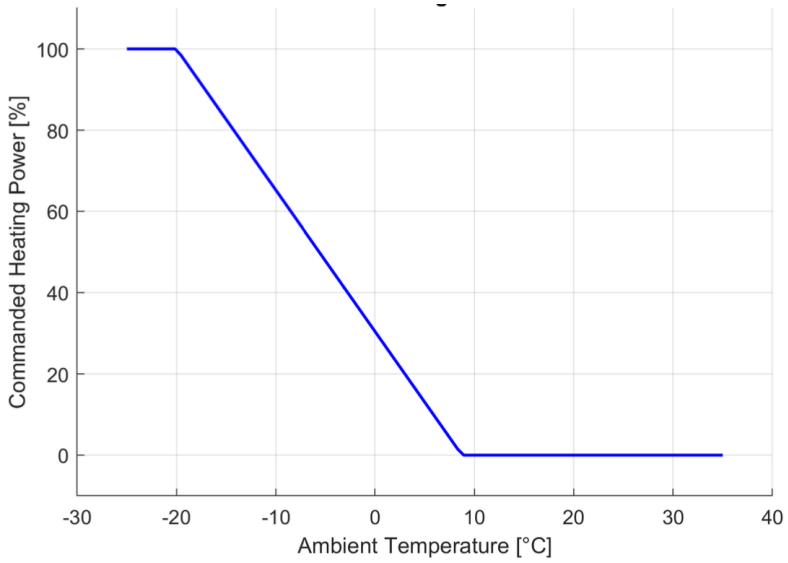


Figure 2.8: Relationship between Ambient Temeprature and Commanded Heating Power percentage for Floor/Wall heating.

2.6.3 Modeling of the Heat Distribution and Heat Flows

It is important to point out that different heating components can influence different heating accumulators. For instance, the HVAC system on the upper deck regulates the temperature of both the upper deck's passenger compartment and the coach body. This occurs because the heated or cooled air circulates through ducts within the coach body before reaching the passenger area. Additionally, the floor and wall heaters impact not only the passenger compartments but also the car bodies they are installed in.
Dead times the are inserted in the model to take into account the inertia of the heat distribution. The example made in the report [14] mentioned that if the heating is activated in the actual towing vehicle, the temperature sensor does not immediately detect an increase

- Dead time for the heat flows coming from the HVAC Unit is equal to 90 seconds
- dead time for the Floor/Wall heating is equal to 300 seconds

External heat Flows to also take into account are:

- Sun Irradiation
- Occupancy
- Chassis Convection

For the scope of this Thesis only the first two will be considered. A better explanation for them is held in the "Methodology" Chapter.

2.6.4 MATLAB Simulink Model of the Thermal Dynamics and HVAC Systems of Half Wagon

Together with the SBB Load Management report [14], a MATLAB Simulink model was developed from scratch to replicate the thermal behavior of half of a wagon considering all of the specifications and assumptions made in the previous sections.

All the parameters, like the k-values, the capacities of the car body and the PID controller variables, had to be estimated.

Subsequently, Dr. Ahmed Aboudonia, worked on the accuracy of the MATLAB Simulink model through fine-tuning, in order to match the available measurement data from Winter 2022 and Summer 2023 with the results from the given Simulink Model.

The overall MATLAB Simulink model is made of four main blocks:

- first block: it selects the reference temperature starting from the ambient temperature, following the piecewise linear function shown in Figure 3.7

- second block: PID Controller (Figure 3.10). The input is the computed reference temperature and the outputs are the different $\dot{Q}$. Specifically three of them are $\dot{Q}_{1,2}$ supplied by the HVAC system and the remaining three by the Floor/Wall heating systems.

- third block: is the heat distribution block. Starting from the $\dot{Q}_{1,2}$ computed in the previous block, it specifies how the heat flows radiate to the chassis, the passenger wagon and the inventory for all of the three decks. It involves a linear transformation.

- fourth block (Figure 3.11): the thermal model block. It takes into account all the disturbances and their non linearities and defines how the temperature in the three decks, which are the outputs that will be fedback in the next step, change.

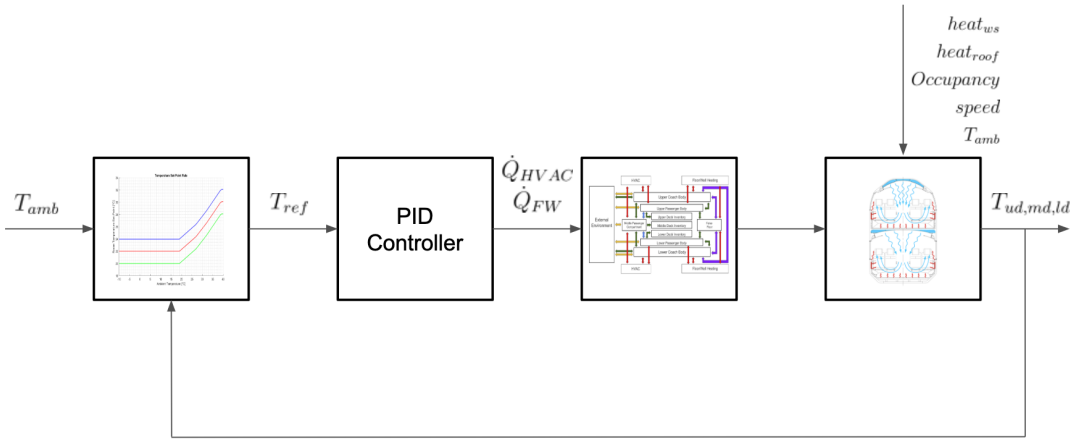


Figure 2.9: The 4 main components of the MATLAB Simulink Model file

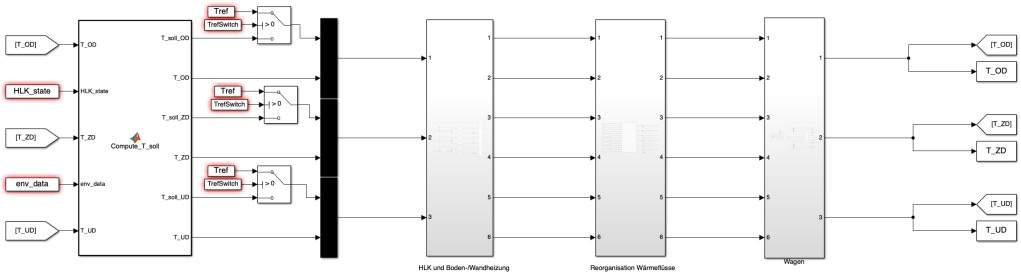


Figure 2.10: Screenshot of the actual MATLAB Simulink Model

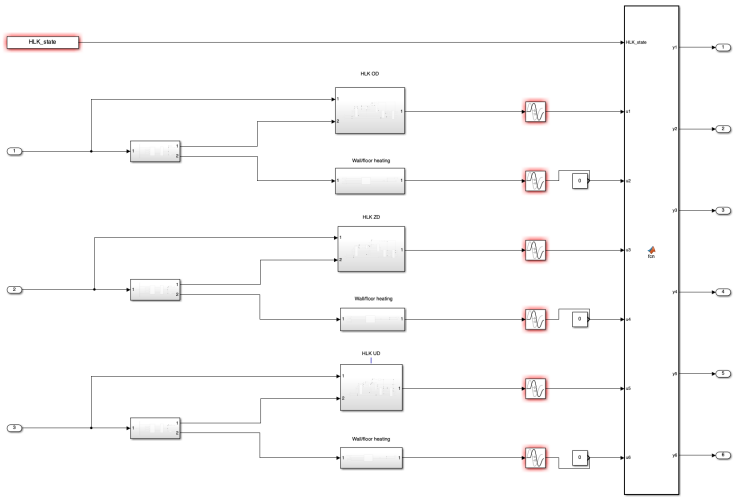


Figure 2.11: Screenshot of the actual HVAC Subsystem of the MATLAB Simulink Model

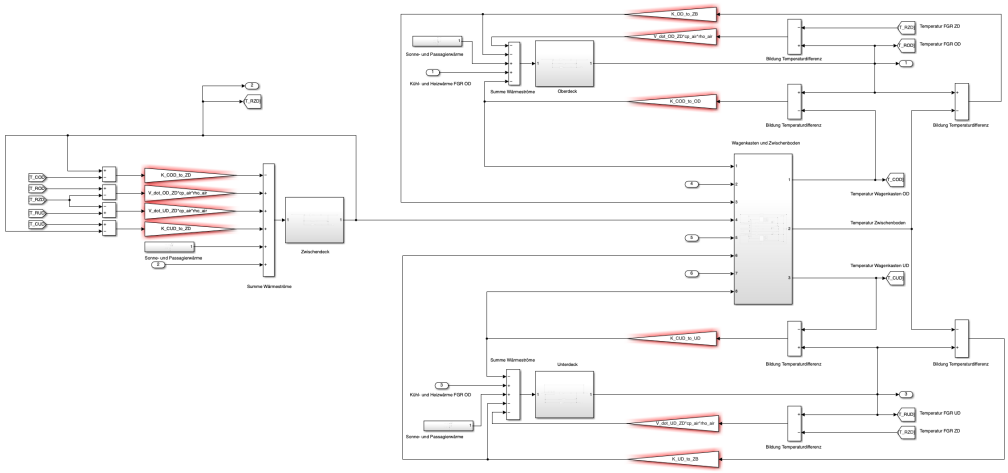


Figure 2.12: Screenshot of the actual Thermal Model of the MATLAB Simulink Model

Chapter 3

Methodology

3.1 Measurement Data

During this Thesis, measurement data coming from the Train Control System (the so called *RDS* from Stadler) which interconnects most of the components of the train, were collected during daily operations. Different Regio-Dosto trains and different days were considered, as shown in the table 4.1. Specifically, it is possible to divide the whole data set into two smaller groups: Winter Data collected in February and March 2022 and Summer Data collected in July 2023.

	Winter Data	Summer Data
Days	22.02.2022 - 28.02.2022 and 11.03.2023 - 31.03.2023	05.07.2023 - 22.07.2023
Total Number of days	28	18
Trains	Regio-Dosto RV 14 Regio-Dosto RV 15 Regio-Dosto RV 16 Regio-Dosto RV 18 Regio-Dosto RV 22 Regio-Dosto RV 23 Regio-Dosto RV 24 Regio-Dosto RV 25	Regio-Dosto RV 14 Regio-Dosto RV 15 Regio-Dosto RV 23 Regio-Dosto RV 24 Regio-Dosto RV 28
Total Number of Trains	8	5

Table 3.1: Train and days from which data were collected.

Data from *RDS* are not the only ones that were used. The following list points out all the sources and all the typologies of data taken into account.

- *RDS* data: as said above, these measurement data come from the Train Control System. The frequency collection varies between the Winter data and Summer data. In the first case, the data are collected every minute, in the second case they are collected every two seconds. They consist of:
  - speed of the Train, measured in [km/h]
  - ambient temperature, measured in [°C]
  - the room temperature divided into the three temperatures of the upped deck, middle deck and lower deck, measured in [°C]

  - the control mode, a boolean value that specifies if the HVAC unit is on (1) or not (0). Originally, also the slumber mode was specified with a different number. For the scope of this Thesis it not of relevance the difference between the off-mode or the slumber mode, so they were merged into off-mode turning this variable into a boolean value

- Weather Data: these data are retrieved from the Meteomatics weather API. The frequency for the Winter Data is every ten minutes and every minute for the Summer Data and consists of:

  - latitude and longitude coordinate of the position of the Train
  - global radiation: is the power per unit area (surface power density) received from the Sun, measured in measured in watts per square meter (W/m^2) in SI units
  - azimuth angle: horizontal angle of the Sun with respect to the North, measured in [°]
  - elevation angle: the angular height of the sun in the sky measured from the horizontal, measured in [°]

- Occupancy Data: a value that goes from 0 to 1 and specifies the occupancy percentage estimated by SBB in the whole train. The heat flow dispersed is estimated to be 75 W per person. It is assumed that the occupancy is constant in every wagon, meaning that an Occupancy value of 0.8 results in 80% of the seats taken in each wagons. The Occupancy may be even greater than 0 in the case of busy trains, with all the seats taken and standing passengers. The estimate of the Occupancy is done using the automatic passenger counting systems, a manual counting system from train conductors and the number of the public transport pass.

(DataLoggers: The Powerloggers we installed can be found here: [PEL103 (chauvin-arnoux.ch)](

3.2 Data preprocessing

Data from RDS, Weather data and Occupancy data all come from different folders, with different structures, different information, different time-zones and, in some cases, different sampling times. The goal is having consistent, standardized and meaningful data structures that can represent the trains setting regardless of the time of the year and the number of the train.
With a Python script using *pandas*, it is possible to put together this different measurements into Excel Files.

3.2.1 Winter Data

Due to the discrepancy in sampling intervals, with meteorological data being recorded every ten minutes and train measurement data every minute, the meteorological data was interpolated to match the train data's frequency. This was achieved by repeating each set of meteorological readings for the corresponding ten-minute period, effectively filling in the gaps. Since the Weather Data do not change so rapidly, this frequency is not a problem.
In the event of missing data, the final Excel files represent these gaps as *NaN* values. This issue arises because not all trains provided continuous data for consecutive days, or due to some data being skipped during the recording process for unknown reasons.

3.2.2 Summer Data

The discrepancy in recording frequency between the train measurements and meteorological data during the summer is notable. Ideally, thirty RDS train measurements are taken every minute (one every two seconds), while Meteomatics data is recorded once per minute. To create consistent Excel files, similar to those used in the winter case, and to enable comprehensive future analysis, an average of the thirty train measurements is calculated.
In the case of missing data, the following assumptions for the train measurement datasets are applied::

- If there are no recordings within a given minute, the data is replaced with a *NaN* value
- If at least one recording is present within the minute, the final value in the Excel file is the average of the available recordings. For example, if only five measurements were taken for the Upper Deck Temperature during a minute, the final value is the average of those five measurements.

Compared to the winter scenario, the summer case has a weather measurement frequency that results in numerous missing data points. Again, to ensure consistency along the datasets and make sure that too many data are discarded, less than ten consecutive weather data are missing, these values are filled with the average between the previous non-missing value and the next non-missing value.

3.2.3 Occupancy Data

Occupancy measurement files differ significantly from the other two types. They include details such as the date, train number, route (comprising departure and arrival stations), departure and arrival times, total number of seats, and the number of occupied seats in both first and second class. Using the latter two pieces of information, the occupancy percentage of the wagon can be calculated and expressed as a value between 0 and 1.
The following assumptions are made:

- the occupancy is constant in every wagon of the train
- only the total occupancy, calculated as total number of seats occupied over the total number of seats, are considered. No difference between first and and second class occupancy is taken into account
- the occupancy is constant during the duration of the route
- in between different routes, if the train stops for less than 20 minutes the occupancy values in that time window is the average between the occupancy of the previous route and the occupancy of the following routes. If the time spent between arrival and next departure exceeds 20 minutes, it is safe to assume that all the passenger left the train in that period and the occupancy is 0.

3.2.4 Additional Parameters Incorporation

In addition to the parameters detailed in Section 4.1, several other relevant parameters are calculated. All of these are computed for every minute of the day, if the parameters from which are computed are missing (meaning they are *NaN* values), they are consequently flagged as *NaN* values.

- temperature reference: as stated in Figure 3.7, It is the target temperature that the deck temperatures should aim to achieve. At the moment, it is calculated based on the ambient temperature.
- bearing: it is the direction of the train with respect to the North when the train is moving. It is calculated starting from a tuple of Latitude and Longitude values retrieved from the Meteomatics data.

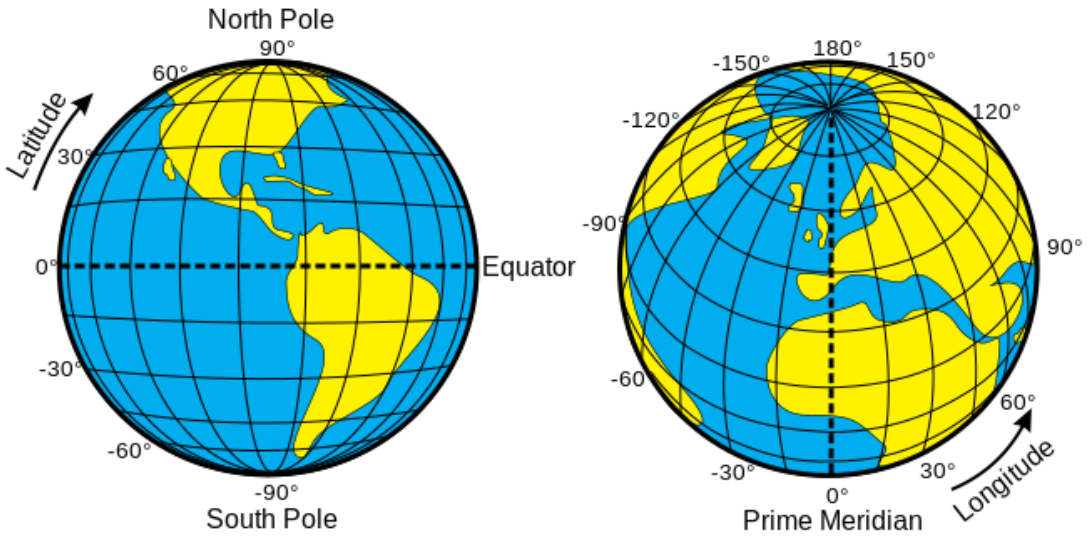


Figure 3.1: Visualization of latitude and longitude on the globe [9]. A position can be uniquely determined by a pair of latitude and longitude values.

Given $lat1$ and $long1$ as the first couple of Latitude and Longitude values in radiant, and $lat2$ and $long2$ as the second couple, the following formula allows to estimate the direction of the train:

$$\text{bearing} = \text{atan2} \cdot \left(\frac{\sin(long2 - long1) \cdot \cos(lat2)}{\cos(lat1) \cdot \sin(lat2) - \sin(lat1) \cdot \cos(lat2) \cdot \cos(long2 - long1)} \right) \tag{3.1}$$

In order to take into account the small measurement errors from the GPS data, the train is considered moving if and if the difference between the two longitude values are greater than 0.001 rad, otherwise the bearing value is equal to zero.

The results are in radiant and can be easily converted to degrees:

$$\text{angle}_{\text{degrees}} = \text{angle}_{\text{radiant}} \cdot \left(\frac{180}{\pi} \right) \tag{3.2}$$

Moreover, to make sure that all the values are positive, it is possible to apply the following formula:

$$\text{angle}_{\text{degrees}} = \frac{\text{angle}_{\text{degrees}} + 360}{360} \tag{3.3}$$

- direction: the absolute direction of the train. It differs from bearing because, if the train is not moving from one minute to the next one, the bearing value is equal zero. The direction parameters takes it into account: if the train is not moving the direction is the last non zero value of the bearing.

- heat flow from the side and from the top: as specified by Hofer in the SBB Report [14], the influence of global radiation was taken into account in the model according to the following equations:

$$\dot{Q}_{sidewall} = \epsilon_{sidewall} \cdot A_{sidewall} \cdot sin(\alpha) \cdot sin(\beta) \cdot \dot{Q}_{irradiation} \tag{3.4}$$

$$\dot{Q}_{window} = \epsilon_{window} \cdot A_{windows} \cdot sin(\alpha) \cdot sin(\beta) \cdot \dot{Q}_{irradiation} \tag{3.5}$$

$$\dot{Q}_{roof} = \epsilon_{roof} \cdot A_{roof} \cdot cos(\alpha) \cdot \dot{Q}_{irradiation} \tag{3.6}$$

Where $\dot{Q}_{irradiation}$ is the Global Radiation presented in section 4.1, $\epsilon_{sidewall}$, ϵ_{window} and ϵ_{roof} are absorption coefficients and $A_{sidewall}$, A_{window} and A_{top} are the side area and roof area in m^2.

The EN 14750-1 standard assumes global radiation values of up to 700 W/m^2 for Switzerland [14].

The overall heat flow from the Sun to the train body is the sum of the previous heat flows computed above:

$$\dot{Q}_{total} = \dot{Q}_{sidewall} + \dot{Q}_{window} + \dot{Q}_{roof} \tag{3.7}$$

α and β are respectively the elevation angle and azimuth angle described in the Weather Data part of the section 4.1, and represented in figure 4.1

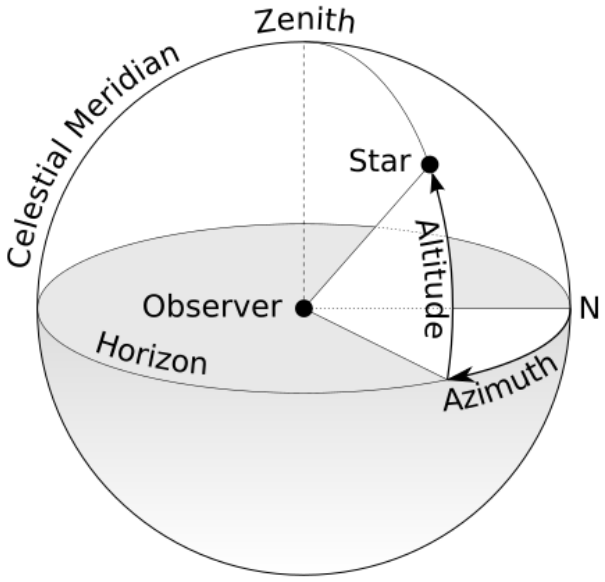


Figure 3.2: Visualization of the azimuth and elevation (here specified as altitude) angles with respect to the North Direction (N) [1]

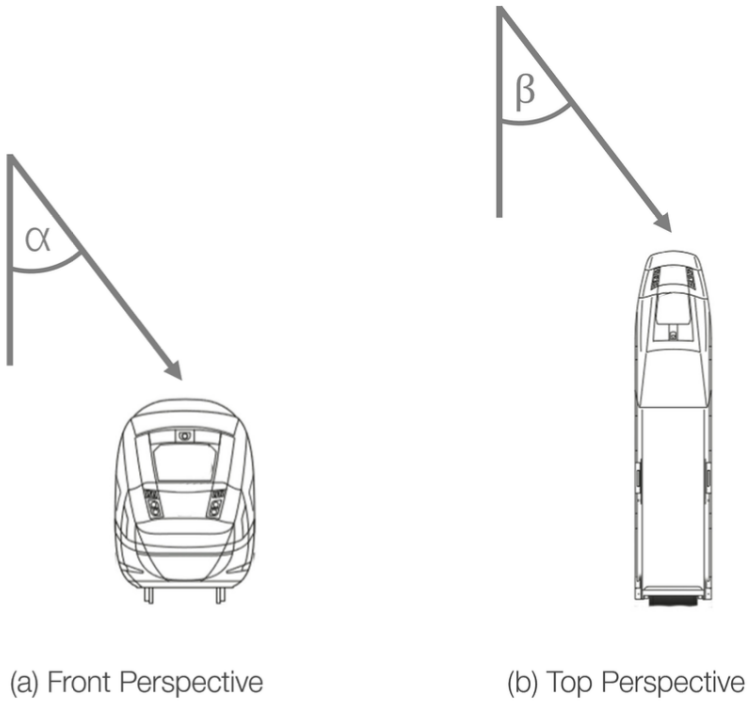


Figure 3.3: Visualization of the relative angles of sun incidence with respect to the train direction [27] [14]

3.2.5 Final Preprocessing Results

As a result, every Excel Files has 1440 rows and 18 columns for each train for each day. Missing data are presented as *NaN* values and only in the case of missing data in every cell the day is not saved.

Every train has its folder named **'RV_0**'** where ** represent the last two digits of the train number. Inside of each folder the Excel file is saved as **'YYYY-MM-DD RV_0**.xlsx'**.

A final recap of all the parameters along with their description and a snapshot of an Excel File is displayed here:

Parameter	Description
datetime	"day-month-year hour:minute" format of the measurement
speed	velocity of the train in km/h
amb_ temp	outside Temperature measured in [°]
ud_ temp	Upper Deck Temperature measured in [°]
md_ temp	Middle Deck Temperature measured in [°]
ld_ temp	Lower Deck Temperature measured in [°]
control_ mode_ HLK1	Whether the HVAC System is in Control Mode (1) or not (0)
t_ ref	Reference Temperature measured in [°]
lat	Latitudinal Coordinate Position of the train
long	Longitudinal Coordinate Position of the train
global_ rad:W	Surface power density received from the Sun measured in $[W/m^2]$
sun_ azimuth:d	horizontal angle of the Sun with respect to the North, measured in [°]
sun_ elevation:d	the angular height of the sun in the sky measured from the horizontal measured in [°]
bearing	Direction of the train with respect to the North if it is moving measured in [°]
direction	Absolute Direction of the train with respect to the North measured in [°]
Occupancy	Occupancy percentage of the train in the range 0-1
heat_ ws	Heat flow received by the train from the side measured in $[W/m^2]$. *ws* stands for windows and side
heat_ roof	Heat flow received by the train from the top measured in $[W/m^2]$

Table 3.2: Description of all the parameters in the Excel Files.

datetime	speed	amb_temp	ud_temp	md_temp	ld_temp	control_mode_HLK1	t_ref	lat	lon	global_rad:W	sun_azimuth:d	sun_elevation:d	bearing	direction	Occupancy	heat_ws	heat_roof
2023-07-05 16:23:00	14,43466667	21,7	24,6	23,5	23,6	1	22,81	47,3877	8,5082	544,7	-360,8665924	42,3	290,8665924	290,8665924	0,364485981	219,0276555	60,60873048
2023-07-05 16:24:00	28,464	21,61666667	24,5	23,4	23,5	1	22,785	47,3885	8,5051	550,6	-355,3542781	42,5	285,1542781	285,1542781	0,364485981	190,278473	48,66406484
2023-07-05 16:25:00	2,698333333	21,6	24,5	23,5	23,4	1	22,78	47,3896	8,4991	568,7	-355,6542781	42,6	0	285,1542781	0,436448598	339,9858566	106,5657743
2023-07-05 16:26:00	0	21,53333333	24,5	23,6	23,4	1	22,76	47,3897	8,4984	573,8	-355,8542781	42,8	0	285,1542781	0,436448598	400,4658846	217,355361
2023-07-05 16:27:00	0	21,5	24,2	23,3	23,4	1	22,75	47,3896	8,4984	574,7	-356,0542781	43	0	285,1542781	0,436448598	415,6673202	319,0236144
2023-07-05 16:28:00	0	21,5	23,8	23,1	23,3	1	22,75	47,3896	8,4984	575,6	-356,3542781	43,1	0	285,1542781	0,436448598	434,1034962	365,7241303
2023-07-05 16:29:00	20,855	21,43	23,6	23	23,2	1	22,729	47,3896	8,4984	576,4	-175,5524148	43,3	104,1524148	104,1524148	0,436448598	133,7269261	447,3591626
2023-07-05 16:30:00	28,93533333	21,4	23,4	22,8	23,1	1	22,72	47,3889	8,5025	550,3	-187,1362166	43,4	115,4362166	115,4362166	0,508411215	295,8935476	459,6114134
2023-07-05 16:31:00	29,57466667	21,4	23,2	22,8	23	1	22,72	47,387	8,5084	500	-182,1643347	43,6	110,2643347	110,2643347	0,508411215	8,957131498	463,905063
2023-07-05 16:32:00	30,41633333	21,4	23	22,7	22,9	1	22,72	47,3855	8,5144	450,3	-187,6347832	43,8	115,4347832	115,4347832	0,508411215	61,90740993	442,8384567
2023-07-05 16:33:00	18,62266667	21,4	22,9	22,6	22,9	1	22,72	47,3836	8,5203	397,3	-186,2952099	43,9	113,8952099	113,8952099	0,508411215	26,39858519	395,9553383

Figure 3.4: Screenshot of an Excel File. The train in question is the RV 15 during the summer period.

3.3 Data Segmentation

In order to use only measurement of interest, it is essential to segment the data based on specific criteria.

- In this Thesis the focus is on the period where the HVAC systems is in automatic mode and fully operational, indicated by the the "**1**" in the *control_ mode_ HLK1* column

- Moreover all the rows with at least one missing value are not considered: these measurement data are removed from the dataset

- Another important criterion it is to segment the data into continuous sequences where no minutes are missing between measurements. This approach guarantees that each data segment contains consecutive data points, with no gaps in the time series. It ensures that, when working with time lags and different time-steps horizons, the data of interest are actually present and accessible

- lastly, another filtering method concerning the occupancy is applied. Since the occupancy values come from an internal estimation from SBB, some pf them are inaccurate. Too many measurement data display the case where no passengers are in the train (Occupancy equal to zero), the speed is greater than 0 km/h and still the HVAC system is in automatic mode: the 24.5% of the total. There may be instances where this accurately reflects reality, such as when a train is returning to another station without passengers or when it is maneuvering within the station's parking area. To exclude very unlikely scenarios, all of the rows which show a a speed bigger than 40 km/h, which is the speed limit within stations and parking areas, and an occupancy equal to 0 are considered incorrect and therefore excluded.

A good example of a data segment is the one in figure 4.4:

- all the eleven rows have no missing values

- all the eleven rows have the *control_mode_HLK1* column with the value "**1**"

- all the eleven rows are consecutive, meaning that there are no gaps between one minute and the next one.

To achieve the segments creation, two Python scripts, **get_winter_days** and **get_summer_days** for the two different periods involved, have been used.

In these original scripts, only segments with more than one consecutive full rows with no missing samples and with the HVAC mode in control mode are taken into account.

Considering all the trains and all the days available, the total number of resulting segments in the Winter and Summer case are presented in the following figure:

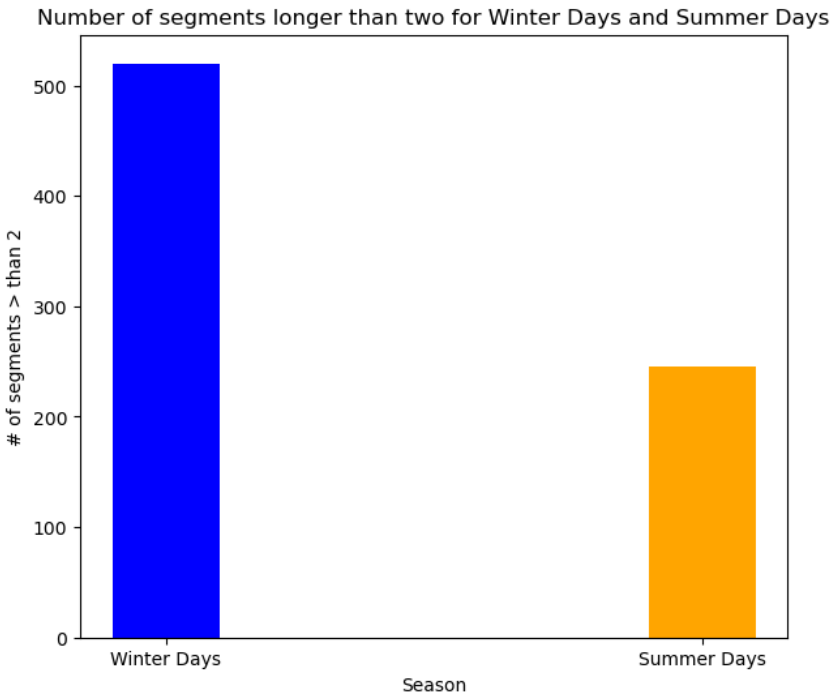


Figure 3.5: Number of segments created as described before, comparison between Winter and Summer Data.

As specified before, the length of the segments will be a crucial element for dealing with different lags and future time steps.
The following figures represent the distribution, from 0 to 100 rows, of the segments based on their lengths and how many segments are greater, respectively, than 30 minutes, 60 minutes and 90 minutes.

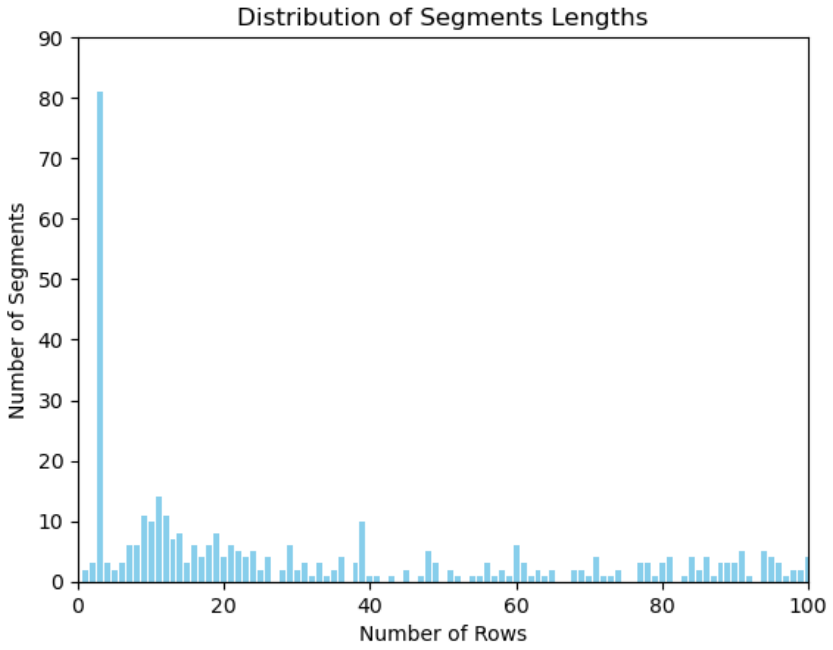


Figure 3.6: Distribution of the segments lengths

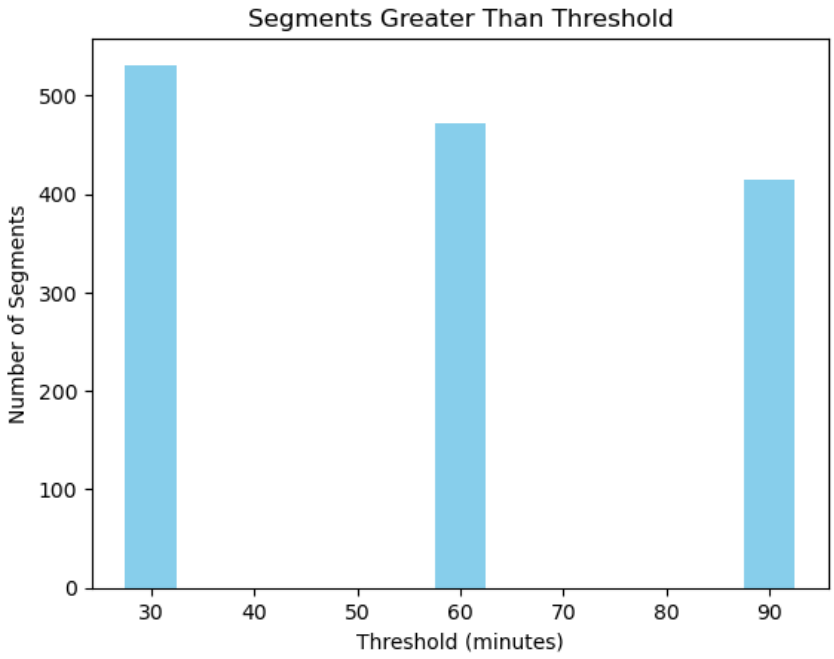


Figure 3.7: Segments greater than 30 minutes, 60 minutes and 90 minutes

The segments are saved as "**segment_***" where * represent the number of the segment starting from from 0. These are saved in specific folders for each of the train, named "**Filtered_RV_0****" where ** is the train number.

3.4 Neural Network-Based Simulator

The first step involving Neural Networks is developing an accurate simulator which can be used in the future to validate new controllers.

This Neural Network-Based simulator is supposed to replace the one developed by SBB using the MATLAB Simulink Model described in the previous chapter.

The goal is trying to predict the three deck temperatures in the following minute, using the parameters of interest showed in the Table 4.2 at the current timestep. In the following section an overview of a Feed Forward Neural Network and some of their most relevant Machine Learning techniques are presented. The actual model is then presented in another section, where all the choices about the fine-tuning results are displayed.

3.4.1 Feed Forward Neural Network (FFNN)

A Feed Forward Neural Network, also called Multi Layer Perceptron (MLP), is the most basic type of Artificial Neural Network. Each neuron in one layer is connected to every neuron in the next layer, resulting in a fully connected network. The strength of the connection between neurons is represented by weights, and the learning process involves updating these weights based on the error of the output.

The goal of an Artificial Neural Network is to approximate some unknown, ideally non-linear, function $f^{(*)}$ that maps an input to an output.

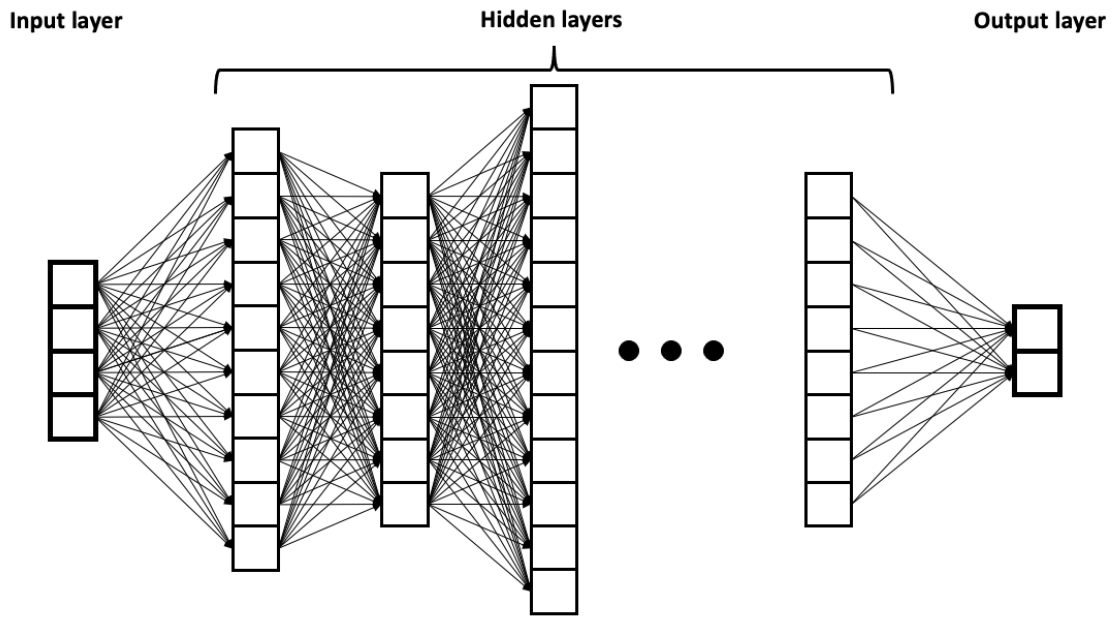


Figure 3.8: Example of a Fully Connected Feed-Forward Neural Network, with an Input layer, and Output Layer and a not defined number of Hidden Layers.

In a FFNN information flows only in one direction (hence, the name Feed-Forward), following the order of the layers in which the neurons are organized: from the input layer, where the data relative to the input x are stored, through the hidden layers, whose number specifies the depth of the Neural Network, until the output layer which specifies the outcome of the overall calculation, which can be a prediction value or a classification value. The number of neurons for each layer represents the so called width of Neural Network.
At each hidden layer, inputs are weighted and summed and in a second step passed through an activation function, introducing non-linearity.
A single neuron at layer l is activated in the following way:

$$z_i^{(l)} = \sum_{j=1}^{n^{(l-1)}} w_{ij}^{(l)} a_j^{(l-1)} + b_i^{(l)} \tag{3.8}$$

where $w_{ij}^{(l)}$ are the weights, $a_j^{(l-1)}$ are the activation values from the previous layer, and $b_i^{(l)}$ is the bias term that shifts the weighted sum.
The passage through an activation function σ can be defined in the following way

$$a_i^{(l)} = \sigma(z_i^{(l)}) \tag{3.9}$$

Among the most famous activation functions there are the $ReLU$ or Rectified Linear Unit, the *Sigmoid Function* and the *Tanh*.
The output at the end of the Neural Network is the following:

$$\hat{y} = a^{(L)} \tag{3.10}$$

Training In the case of *Supervised Learning*, where the datasets are complete and consist of pairs of inputs and outputs, the training process involves comparing the predicted output after the Feed-Forward pass described above and the actual output.

In order to train the neural network, a loss function has to be minimized. For regression tasks, usually the mean squared error (MSE) is used.

$$\mathcal{L}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3.11}$$

or the cross-entropy loss for classification tasks:

$$\mathcal{L}(\mathbf{y}, \hat{\mathbf{y}}) = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right] \tag{3.12}$$

Once the error is computed during the feed forward pass, the backpropagation algorithm is used to update the weights and biases by computing the gradients and applying gradient descent:

$$w_{ij}^{(l)} := w_{ij}^{(l)} - \eta\frac{\partial\mathcal{L}}{\partial w_{ij}^{(l)}} \tag{3.13}$$

$$b_{i}^{(l)} := b_{i}^{(l)} - \eta\frac{\partial\mathcal{L}}{\partial b_{i}^{(l)}} \tag{3.14}$$

where η is the learning rate. Usually, instead of using all the data points at once to calculate the new weights and biases through the gradient descent algorithm, a batch of data points is used to compute the gradient and then used to update the model parameters.
Once all the batches, and so all the data, are used it is said that an *epoch* has passed.

The Risk of Overfitting Neural Networks learn patterns from data and can then use these patterns to make future predictions on unseen data. Relying too much on the available data for the training process can lead to the so called *Overfitting* problem: the model results to be excessively complex, capturing the random fluctuations in the training data, resulting in poor generalization.
The main reason for this to happen are:

- a too high number of parameters which allows a too precise fitting to the training data
- a small dataset
- poor variation in the training dataset which leads the Neural Network to have a poor capability to handle different inputs

In order to prevent overfitting, the main techniques are:

- dividing the dataset into training and testing data: only the training data are used to learn patterns, the remaining data are treated as unseen data with which testing the Neural Network. A classical split between training and testing data is 80-20% depending on the dimension of the dataset
- regularization techniques: adding a penalty term to the cost function to prevent complex models to be chosen
- dropout regularization: drop the use of some random neurons in the layers to prevent
- Early Stopping: check the model performance on a testing set and stop training when the testing performance is not meaningful anymore.

3.4.2 The Simulator Model

In the case of the simulator, the objective is to predict the temperatures of the three decks (Upper Deck, Middle Deck, and Lower Deck) for the next minute, based on the parameters known at the current timestep. In Machine Learning language, the *output features* are then the three temperatures at the next timestep and the *input features* are the chosen parameters.
The architecture of choice is a Feed Forward Neural Network.
The following pictures shows which input and output features are needed.

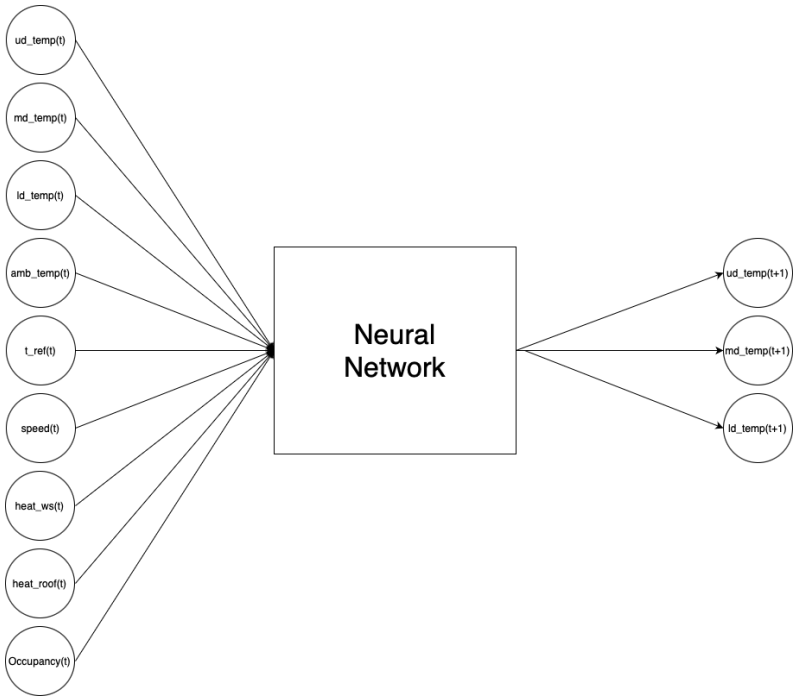


Figure 3.9: Representation input and output labels for the Simulator Model. The Neural Network here is presented only as a black-box model, with no information about the number of layers and number of neurons.

The number of lags for a prediction task represents the number of previous time steps whose values are retained and used as input features for forecasting future values. This essentially means how far back in time the model looks to gather the data needed for making accurate predictions.
In this case the number of lags is equal to zero, meaning that only the parameters for the current timestep are retained and used.In terms of the number of samples and segments, all segments that are at least two consecutive minutes long can be utilized. This results in a total of 177,000 data points, derived from all trains and all days, being available.

NN Architecture The number of neurons and layers is carefully chosen to match the complexity and typology of the task. In this case a Feed Forward Neural Network with 3 hidden layers, each of them with 10 neurons resulted to be an optimal choice.
The input layer, as already shown in the Figure 3.9, has 9 neurons representing the input features, while, since the goal is the prediction of the three decks temperatures at the next timestep, an output layer of three neurons is needed.

Training Process and Fine tuning It is of key importance to get an accurate prediction without excessively overfitting the training data. Because of this, a 80-20% split of the training

data is used to facilitate an objective evaluation of the chosen architecture.
Other fine tuning techniques are the following:

- the training and test data are shuffled, allowing the big dataset to be more mixed and not ordered by time, avoiding the risk of learning patterns that are due to the training and evaluation separation
- an early stopping function is deployed to avoid overfitting: the number of epochs is then not very relevant because, in all likelihood, the training procedure will be stopped due to this function
- weights and biases are initialized with *Xavier Initialization*: is to initialize the weights such that the variance of the activations are equal across each layer. This method helps prevent the gradient descent from exploding or vanishing leading to a better optimization
- the datapoints are divided into small batches made of 32 points and, instead of using the normal classic descent algorithm, a stochastic gradient descent one is used
- the optimizer is the Adam optimizer, which stands for Adaptive Moment Estimation: it is an adaptive learning rate algorithm. It dynamically and adaptively adjusts the learning rate for each individual parameter within a model, instead of using a global learning rate
- the training loss is the Mean Squared Error for training and Mean Absolute Error for evaluation. The formulas will be better discussed in the Result Chapter
- ReLu function are employed as activation functions
- both the training and evaluation records are tracked and their behaviors are represented in diagrams.

Scaling To help the training process with a consistent gradient descent, the input data were scaled in order to match their magnitude. Since the output of the Neural Network are the three decks temperature and, among the inputs, other temperature labels are considered, such as ambient temperature and reference temperature, the other labels were scaled to match the order of magnitude of 10, namely the magnitude of most of the temperature in the dataset.
To this regard:

- Occupancy values were multiplied by a factor of 10. This does not match the order of magnitude of 10, but it allows to get a value closer to the temperature ones
- speed values were divided by 10
- heat flows values were divided by 100

3.5 Neural Network-Based Temperature Prediction with Multiple Lags and Time Horizons

The case of the simulator is a very isolated one: starting only from the current timestep information the goal is to predict as best as possible the temperatures of the three decks in the next minute.
As said before, this is the case of 0 lags, where there is no need to go back in time and check for past data to make a prediction.
In this way though, the behavior of the overall system is not taken into account: the use of lags

allows to go back in time and retrieve more information of how the wagon temperatures changed with time and it is possible to find patterns that enhance the prediction performance.

Of course, the number of datapoints that can be emplyed for this scope are not the same ones as for the simulator. If a time lag of 10 minutes is used to go back in data history to retrieve past information and predict the (t+1) temperatures: only the segments which are 12 rows long can be utilized for training, the past 10 rows, the current one and next one which gives the actual values to be predicted and is used for supervision in the context of supervised learning.

Moreover, not only the number of lags can vary, but also the number of steps in the future that are to be predicted. For example, unlike the Simulator case, with a time horizon greater than 1, it is possible to predict the temperatures in 2, 5, 10 minutes and so, with, of course, a worse accuracy as the time horizon gets greater and greater.
Recovering these information and using them in a specific Neural Network is fundamental to get a better grasp of the underlying functioning of the thermal model of the wagon: this will be done using not only a NN techniques, but implementing them with some physical equations that rule the thermodynamics behavior.

3.6 Neural Network-Based Thermal Model: PCNN

The second step involving Neural Networks concerns the creation of a specific type of Physics-informed Neural Network, or *PiNNs*, called Physically Consistent Neural Network, *PCNNs* [17]. Physics-informed Neural Networks take into account underlying physical laws that rule the system of interest and use this prior knowledge to affect the Neural Network behavior, and avoid the classic generalization issues typical of Machine Learning methods.

3.6.1 Physics-Based models, Black-box models and Hybrid models

Physics-Based models Physics-Based models make use of Physics laws, typically Ordinary Differential Equations to mimic the functioning of the considered model. In the case of HVAC Systems and wagons thermal model, the physics behind can be the one ruling the thermal dynamics of the environment through convection, conduction and radiation equations.

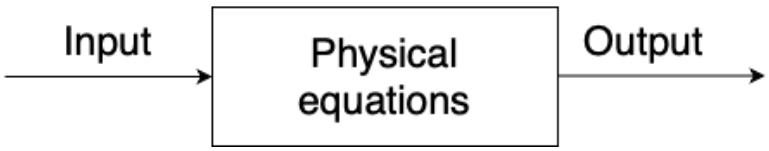


Figure 3.10: Representation of a White-Box Model.

These thermal models, though, are very difficult to mimic and a lot of assumptions and simplifications, which limit their accuracy, have to made. An example of assumption in the context of the HVAC System and the train wagon is the one regarding the half-wagon behavior: it is assumed, indeed, that the two halves behave exactly the same, each part being only controlled by its HVAC System. And while it might sound like a small assumptions, it is very unlikely in practice to imagine that the two halves have an identical functioning and that do not interfere with each other. Developing an accurate model, in this case, requires a lot of time, the calibration, the parameters implementation and the cost are then relevant.

Black-box models In contrast to the Physics-based models, black box models completely rely on the data and computational resources to identify patterns between the inputs and the outputs.
Example of Black-box models are Neural Networks, Support Vector Machines or Random Forests: in all these cases the internal functioning, even if carefully engineered, is not clear and interpretable. Hence, the term "black-box," which evokes the idea of something occurring behind the scenes, or data-driven to highlight the importance of the data.

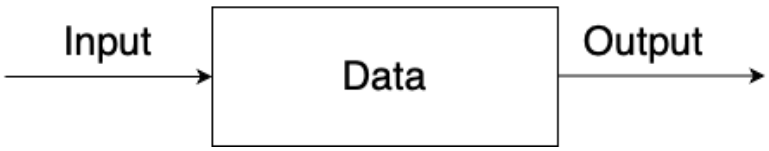


Figure 3.11: Representation of a Black-Box Model.

In this scenario, precise physics equations are not needed, unlike in the Physics-Based models, but previous data and computational power are essential for Black-box models. Even if the succeed in capturing the non-linear dynamics intrinsic to the systems, there is no guarantee that they can actually follow the underlying physical laws and the risk of overfitting might affect the general outcome.

Hybrid models A middle ground between the two aforementioned models are Hybrid models, which make use of both of physics laws and past data: because of this reason they are called Gray-box models. Usually, Gray-box models start with linear state-space models and then, with the help of typical techniques from black-box models, try to learn from the data the non-linearities that can be present.
An example of Hybrid Model is, as mentioned before, physically informed Neural Networks.

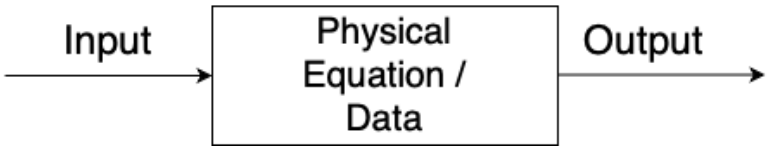


Figure 3.12: Representation of a Gray-Box Model.

PCNNs The reference used in this Thesis for the Physically Consistent Neural Networks comes from the paper "Physically Consistent Neural Networks for building thermal modeling: Theory and analysis" [17].
In their work, L. Di Natale et al., presented a new PiNN architecture which can preserve the same good accuracy of classical Neural Networks, but still be consistent with the physical laws that rule the thermal behavior of the building.
The term "physically consistent" refers to the fact that any change in the input will cause a change in the output, and this effect follows physical laws.
Although having a constrained structure due to the integration of the physical principles, it was possible to show that the PCNN have a less tendency to overfit the data.
In particular, the inputs are simultaneously processed by two modules: one to preserve physical exactness when feasible, and the other to capture nonlinearities using a shallow neural network, so what it's not easy to characterize a priori.

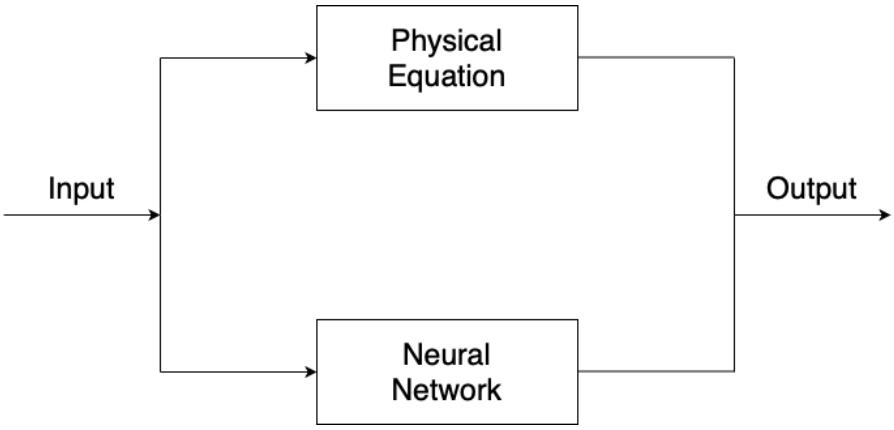


Figure 3.13: Basic representation of a PCNN Architecture.

All the parameters, from both of the modules, are optimized in one single step.
Keeping in mind that the theory of comfort for public transport derives from building comfort theory, it is fair to assume that this particular type of Gray-Box model can achieve interesting results also for the passenger wagon in train like the SBB Regio-Dosto.

3.6.2 PCNN Model for Temperatures Prediction in a SBB Regio-Dosto Train

The figure 3.13 represents a general and basic architecture for a Physically Consistent Neural Network: in the specific case of the temperatures prediction for half of the wagon of the SBB Regio-Dosto it is important to specify further elucidations.
The overall goal is to use this specific model as an input for a Model Predictive Control problem: generally, Artificial Neural Networks do not lead to convex input-output relationships, making them not ideal candidate for an optimization problem as the one in a regular MPC [5].
The main idea here is using the set of inputs in a separated way to preserve convexity and using the PCNN architecture for the MPC formulation.

Among all of the various inputs features, the ones of the current timestep in consideration, the ones of the previous lags and the ones of the forecast for the following time horizons, some of them can only be used in the Physical Equations part of the architecture. The other input features all contribute to model only the non-linearities that affect the system: the velocity of the train, the various heating depending on the direction and the position of the train, the Occupancy
This expedient guarantees that the the relationship between inputs and outputs is convex and usable within the MPC scope [5] [17].

The general prediction formula can then be summed up like this:

$$x(t+1) = Ax + Bu + \rho(x) \tag{3.15}$$

in which the prediction is made of two different components, a linear part depending on the current states and the inputs, and a non linear part to define the non linearities: ρ.
For the case of the half wagon of SBB Regio-Dosto, as seen in [14], the physics equations used to describe this model derive from the RC model.
The equation to choose has to take into account the consistency of the thermal dynamics, and since the main gradient that drives the heat exchange is defined by the difference between the temperature of the decks and the ambient temperature, the linear relationship in the case, for example, of the one step prediction from (t) to $(t+1)$ and zero lags is:

$$\begin{pmatrix} \text{ud\_temp(t+1)} \\ \text{md\_temp(t+1)} \\ \text{ld\_temp(t+1)} \end{pmatrix} = a \cdot \begin{pmatrix} \text{ud\_temp(t) - t\_amb(t)} \\ \text{md\_temp(t) - t\_amb(t)} \\ \text{ld\_temp(t) - t\_amb(t)} \end{pmatrix} + b \cdot \text{t\_ref(t)} \tag{3.16}$$

This is only one of the many option that can represent the underlying physical equations that rule this thermal system. The main thing, to preserve the convexity aforementioned, is to have the three decks temperatures (ud_temp, md_temp, ld_temp) and the reference temperature (t_ref) only in the linear part and not as inputs of the NN. This is because (t_ref), in the second step of the MPC calculation, is the optimization parameter.
In the equation 3.16, for the case of one step prediction, the matrix a is a 3 by 3 while b is a 3 by 1. With the increasing number of timestep horizons these matrix get bigger and bigger as well. At the same time, also the architecture of the Neural Network has to be more complex because of the growing number of parameters if a longer horizon is chosen.

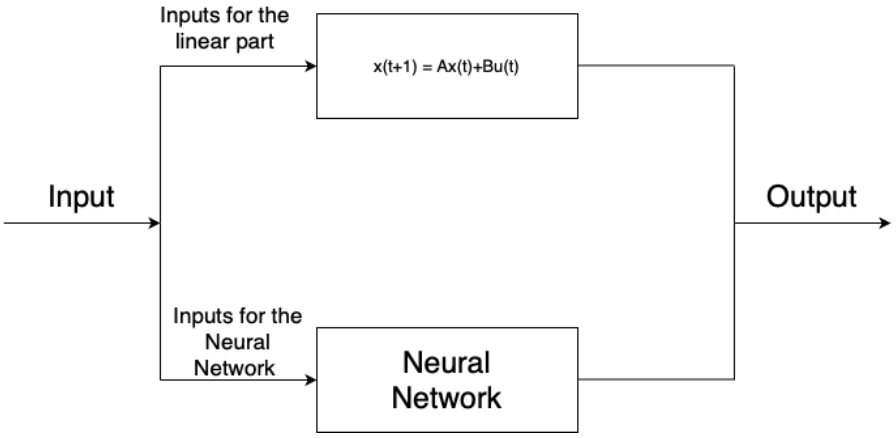


Figure 3.14: PCNN Architecture with a linear module for the physical equations and the separation for the inputs.

The case with different lags and horizons and the splits between the inputs are described in the Results and Discussion Chapter.
As for the fine tuning techniques, the same consideration of the subsection 3.4.2 about the Simulator hold.

Chapter 4

Results and Discussion

4.1 Simulator

Using the architecture and fine tuning techniques described in the previous chapter, the results of the Feed Forward Neural Network Simulator are presented in the following picture:

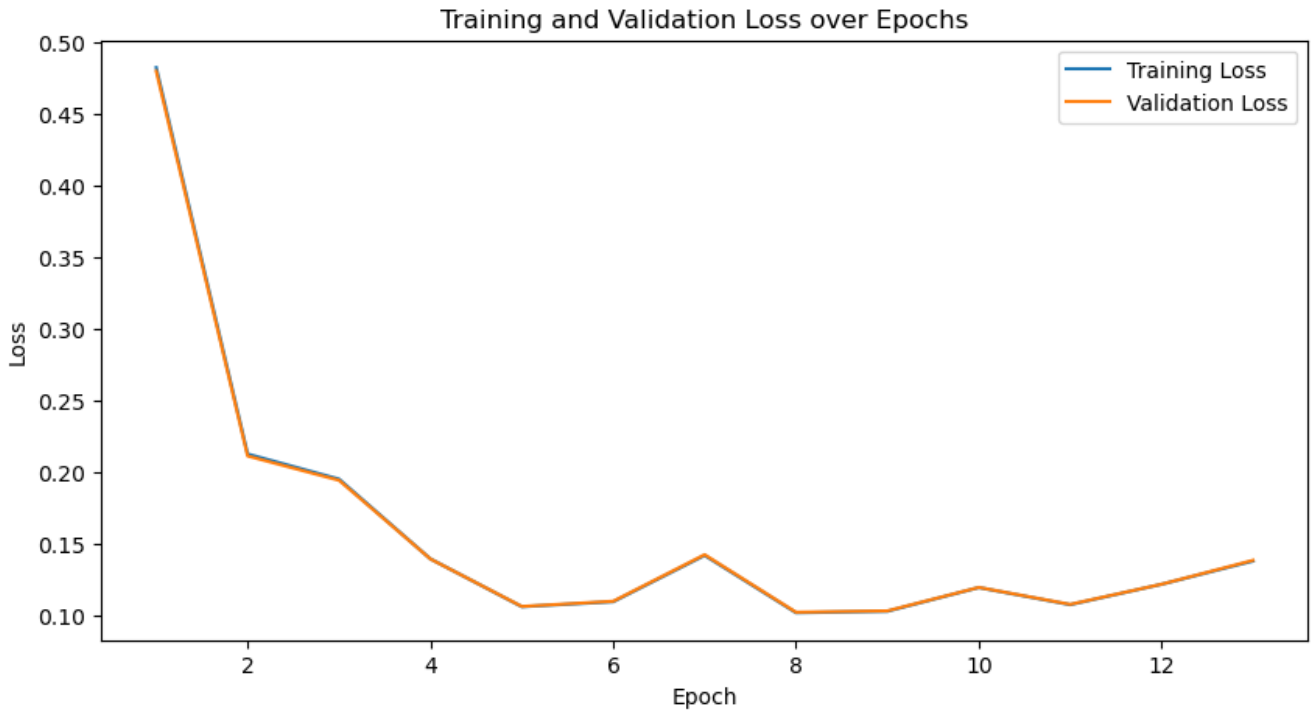


Figure 4.1: Training and Evaluation error for the FFNN Simulator

On the x-axis the number of epochs are shown, while the Training Loss and Evaluation Loss are represented on the y axis.

The input features, once again, are:

- the three decks temperatures
- the ambient temperature
- the reference temperature

- the speed
- the heat exchanges due to irradiation, on the side and on the roof
- the occupancy percentage

The first noteworthy thing is that the two losses almost exactly overlap. This is due to the fact that dataset itself is not that diverse: both the training and evaluation datasets are very similar, regardless of the fact that a shuffle option was employed. In most of the cases the temperature at the next minute is the same, or very similar at least, of the current one.
Because of the non-variety of the overall dataset, the challenge was choosing the right architecture without overfitting too much.
Only a few number of epochs are needed to let the Early Stopping method to stop the training procedure in advance: already after 5 epochs the error between the real temperature and the prediction is just 0.1 °C.
Using a different training/testing split produces a similar result in terms of overlapping but a worse result in terms of error, showing that the rule of thumb of a 80-20% split yields a better result.

Figure 4.2: Training and Evaluation error for the FFNN Simulator, 50-50% split between training and testing error

While the Mean Squared Error (MSE) formula for the training function cost to be minimized was already presented in previous chapter, here the MAE is used as conveys the result in a better way: it is the Mean Absolute Error between the predicted value and the actual value on the average for all the datapoints.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{4.1}$$

where n is the number of data points, y_i is the actual value of the temperatures and $\hat{y}_i$.

The result of 0.1 °C as MAE for the simulator is definitely a good one: it is important to keep in mind that the sensitivity of the sensors used for data recording is, in fact, 0.1 °C.

4.2 PCNN used as a Simulator: a comparison

Even if not used for this scope, it is possible to compare the results between the Normal FFNN Simulator with only the neural network component and the PCNN. In this case the Neural Architecture is the same and only the added linear module changes.
The input for the linear module are:

- the three decks temperatures
- the ambient temperature
- the reference temperature

used as expressed in the equation 3.16.

The remaining input features are employed as inputs for the Neural Network part, with the exception of the ambient temperature which is still used also for the FFNN.

A few more clarifications which need to be done in the case of the PCNN:

- the initialization of the parameter a and b it is of crucial importance: since, as the name says, it has to be Physically consistent, these parameter need to be initialized in a meaningful way. Many attempts were done in this regard and the best way of optimization turned out to be setting them equal to Identity matrix for a and to the vector of ones for b. The other parameters, weights and biases of the FFNN, are still used initialized through *Xavier Initialization*
- Forecast prediction: it is assumed that all the future disturbances like speed, occupancy and weather which affects the heat exchange due to irradiation are perfectly known since coming from past data retrieved by SBB and Meteomatics. In the case of a real experiment in a non simulation environment, this might not be the case and these data could be less reliable depending on the time window.

Figure 4.3: PCNN used as a Simulator

The number of epochs and error are very similar to the ones of the original simulator, mainly because the task is still the same, the are not many more information given to this architecture and the same argument about the variability of the data still holds.
The real difference is that, during the same optimization step, the parameters a and b are optimized and it is interesting to see how they change in this case, starting from the initialization values imposed by the physically consistency.
At the end of the 8 epochs,

$$a = \begin{pmatrix} 0.969 & 4.6592e(-3) & -9.8573(e-3) \\ -5.3441e(-3) & 0.993 & -1.6058e(-2) \\ 2.0970e(-3) & 8.5291e(-4) & 0.971 \end{pmatrix}$$

$$b = \begin{pmatrix} 0.5129 \\ 0.5325 \\ 0.5313 \end{pmatrix}$$

showing that the temperatures at the next timestep strongly depend on the ones at the current timestep (the matrix is still very close to the identity one), and each deck influences each other very slightly, two magnitude orders less.

Trying to use the same architecture, with same inputs but to predict the temperature in the next three minutes yields this result:

Figure 4.4: PCNN used as a Simulator to predict the decks temperatures in after three minutes

In this case the accuracy is smaller with a MAE of 0.2°C and the correlation between the temperatures at $(t+3)$ and temperature at t are different:

$$a = \begin{pmatrix} 0.891 & 0.016 & -0.015 \\ -0.043 & 0.98 & -0.046) \\ -0.01 & 0.008774 & 0.92 \end{pmatrix}$$

$$b = \begin{pmatrix} 0.539 \\ 0.531 \\ 0.5179 \end{pmatrix}$$

Here the values on the diagonal of the matrix a are getting further and further from 1, while the other ones get bigger compared to previous a. Some of these are also with negative values: it was not imposed to them be strictly positive or greater than zero, but this could be something for the next dissertation regarding the topic.

4.3 PCNN with Lags and Time Horizons

The last section of this chapter is used to discuss the results of the PCNN when a predefined number of lags and time horizons is set.
In this case the number of parameters can vary a lot:

- for each minutes of the future horizon, 5 new disturbances need to be added: speed, occupancy, the two heat flows and ambient temperature
- for each minute of the lags in the future, 9 new disturbances are added: speed, occupancy, the two heat flows, ambient temperature, decks temperatures and reference temperature. In

these case, past references temperatures and decks temperatures can be taken into account, since are not going to be matter of optimization during the MPC.

It is important to highlight, in this case, how many different combination of inputs to this architecture can be produced: in fact, past ambient temperatures, for example, can be used with no distinction as FFNN input or Linear Module input. The crucial point is that current temperatures and reference temperatures are only inputs to the linear part.
In this case, it was decided to keep the a similar separation as before:

- Linear inputs: decks temperatures, ambient temperature and reference temperature both from the current timestep and the past one are inputs to the linear Module
- speed, ambient temperature, heat flows and Occupancy of the current timestep, the past ones and the future ones as inputs to FFNN

Due to these choices the a and b matrices get fatter as more and more new matrices base on the number of the past lags.
Depending on the number of lags and the horizon window, different combinations are possible. Since the number of parameters increase, also the complexity of the FFNN architecture needs to be adjusted and optimized.
With a number of lags equal to 10 and a time horizon equal to 10, the neural network architecture is set to be ideal with 10 hidden layers and 16 neurons.
The result is the following:

Figure 4.5: PCNN, Lags = 10, Horizon = 10

Chapter 5

Conclusion

In this thesis, the focus was on the feasibility of reproducing a correct prediction architecture using a Feed Forward Neural Network, in the first place alone, and in the second place in parallel with a linear module which contains an exemplification of the physical laws that rule the system of the half wagon of a SBB train. The data used to train the model were given directly from the sensor recording of SBB over the period of Winter weeks and Summer Weeks.
The study was conducted in a simulative context, in preparation for the potential Model Predictive Control that can be derived starting from the results of this study.

Firstly, a Simulator was developed: the main goal of it is, through a Feed Forward Neural Network, predicting the temperatures at the next timestep based only on the data of the current time, with no data history retrieved. This simulator was able to reach an accuracy for the Mean Absolute Error of 0.1°C.
Then, a novel type of Physics Informed Neural Network (PiNN), named Physically Consistent Neural Network is employed to impose convexity for input-output mappings to guarantee that the problem can be used within the MPC framework.
An example of a PCNN with 10 lags and 10 time horizons is then shown to prove the validity of the PCNN.
Many assumptions were made during these work, in particular no further restrictions on the linear parameters were imposed, meaning that more research in that regard is needed. The weather forecast, which is gonna play a key role in the MPC pipeline, was also supposed to be perfectly known, which is not true in reality since only estimate of the future forecast can be retrieved in real time.
Moreover, only the most basic typology of Neural Network, meaning the Feed Forward one, was used to to model for the first time the thermal dynamics of the passenger carriage. Many different NN model could be examined in the future: Recurrent Neural Network (RNN) and Long-Short Term Memory Neural network (LSTM) were proven to be very effective in the prediction of temperature in buildings taking into account the past lags [17].

Bibliography

[1] 2020.

[2] Accessed: 2024-06-10. Feb. 2023.

[3] Dr. Ahmed ABoudonia. *Train Climate Control, ETH/SBB Meeting Powerpoint Presentation.* Feb. 2023.

[4] Bunning et al. "Experimental demonstration of data predictive control for energy optimization and thermal comfort in buildings". In: *Energy and Buildings, Volume 211,109792,* (2020).

[5] Bunning et al. "Input Convex Neural Networks for Building MPC". In: (2021).

[6] Bunning et al. "Physics-informed linear regression is competitive with two Machine Learning methods in residential building MPC". In: *Applied Energy,Volume 310,118491* (2022).

[7] Marco Molinari Giorgio Pattarello Luca Fabietti Karl H. Johansson Alessandra Parisio Damiano Varagnolo. "Implementation of a Scenario-based MPC for HVAC Systems: an Experimental Case Study". In: *IFAC Proceedings Volumes, Volume 47, Issue 3,Pages 599-605* (2014).

[8] Bernd Claus. *Benutzerhandbuch Klimaanlage S49T.* Internal Document Stadler, SBB Faiveley. 2016.

[9] Djexplo. *Own work, CC0.* May 2011.

[10] *EN 14750-1:2006.* European Standards from CEN, CENELEC and ETSI. 2006.

[11] Swiss Federal Office of Energy. *SBB Electrical Power.* Accessed: 2024-06-06. Aug. 2021.

[12] Tullio de Rubeis Dario Ambrosini Alessandro DâInnocenzo Rahul Mangharam Francesco Smarra Achin Jain. "Data-driven model predictive control using random forests for building energy optimization and climate control". In: *Applied Energy,* (2018).

[13] Urs-Peter Menti Sina BÃ¼ttner Franz Sidler Mathias Niffeler. *SYNTHESEBERICHT ENERGIEEFFIZIENZ IM BEREICH HEIZUNG,LÃFTUNG, KLIMA UND FAHRZEUGHÃLLE IM ÃFFENTLICHEN VERKEHR.* Accessed: 2024-06-06. Feb. 2021.

[14] Jan Gasser. *Machbarkeitsstudie zu den dezentralen Aspekten des prÃ¤diktiven thermischen Lastmanagements, SBB Report.* Automatic Control Laboratory, Swiss Federal Institute of Technology Zurich. 2019.

[15] Matthias Tuchschmid Johannes Estermann. *Internal Powerpoint Presentation: Projektskizze Â«prÃ¤diktives Heizen.* Accessed: 2024-06-06. Nov. 2021.

[16] I. Baier F. KÃ¤ser. *Konzept Klimaanlage Dosto.* Internal Document Stadler SBB. 2009.

[17] P. Heer C.N. Jones L. Di Natale B. Svetozarevic. "Physically Consistent Neural Networks for building thermal modeling: Theory and analysis". In: *Applied Energy* (2022).

[18] IEA Building Energy Performance Metrics. "Supporting energy efficiency progress in major economies". In: *IEA: Paris, France* (2015).

[19] EPFL Prof. Francesco Borrelli UC Berkeley Prof. Melanie Zeilinger Prof. Manfred Morari University of Pennsylvania Prof. Colin Jones. *Model Predictive Control, ETHZ Course*. 2021.

[20] *RABe 511 Lastenheft MPC fÃ¼r HLK*. Internal SBB Document. Apr. 2023.

[21] *Regio/IR-Dosto (SBB RABe 511)*. Available: https://www.aufdenschienen.ch/Regio-IR-Dosto/Regio-IR-Dosto.htm. Accessed: 2024-06-11. 2020.

[22] SBB. *Regio double-decker*. Available: https://www.sbb.ch/en/travel-information/services-on-train/our-trains/regio-double-decker.html Accessed: 2024-06-11.

[23] *SBB Electrical Power*. Accessed: 2024-06-06, Available: https://company.sbb.ch/en/sbb-as-business-partner/services-rus/energy.html. 2024.

[24] *SBB Facts and Figures*. Accessed: 2024-06-06, Available: https://reporting.sbb.ch. 2024.

[25] Michael JoÃ«l Schnetzler. *MPC for a Train HVAC System*. Automatic Control Laboratory, Swiss Federal Institute of Technology Zurich. 2022.

[26] Nico Wertli. *6th WorkshopPredictive HVAC control*. SBB Westlink Building, Meeting and presentation. Apr. 2024.

[27] Nico Wertli. *Data-driven Learning and Optimization-based Control for HVAC Systems in SBB Trains*. Automatic Control Laboratory, Swiss Federal Institute of Technology Zurich. 2023.

[28] Wikipedia. *Stadler KISS*. Available: https://en.wikipedia.org/wiki/Stadler$_K ISS Accessed$: $2024 - 06 - 11$.

Appendix A

Chapter

A.1 Section

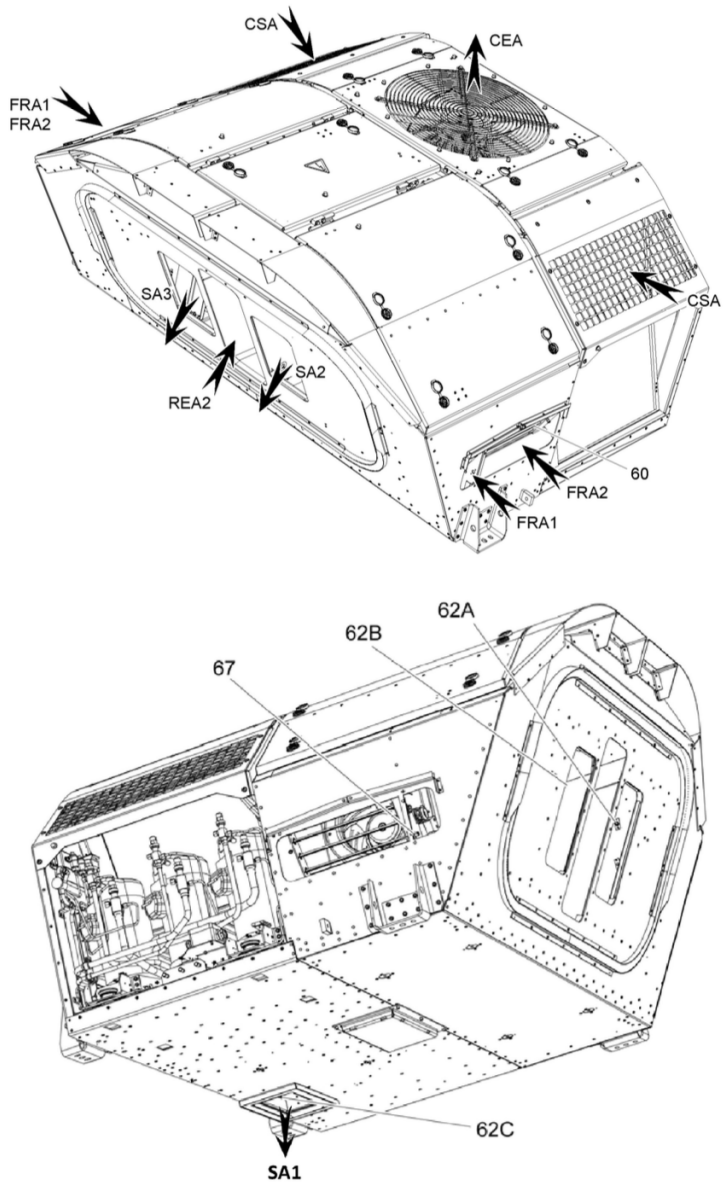


Figure A.1: Drawing of the HVAC System [8]

HVAC Output Limits [W]	Upper Floor	Middle Floor	Lower Floor
Max Cooling Power	-10'000	-8'000	-10'000
Max Heating Power	8'000	5'000	10'000

Table A.1: Physical limitations of the HVAC unit [8]

FW Heatings Output Limits [W]	Upper Floor	Middle Floor	Lower Floor
Max Cooling Power	0	0	0
Max Heating Power	4'994	2'710	4'021

Figure A.2: Table representing the Outside Air Intake Volumetric Flow Rates [14]

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of Originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of Master's thesis (in block letters):

Authored by (in block letters):
For papers written by groups the names of all authors are required.

Name(s):	**First name(s):**

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date	**Signature(s)**

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.