UNIVERSITÀ
DEGLI STUDI
FIRENZE

Doctoral Programme in Industrial Engineering
Dottorato di Ricerca in Ingegneria Industriale

XXXV

Decarbonisation and Machine Learning-based energy monitoring of Natural Gas City Gate Stations

ING/IND-09

Doctoral Candidate

Ing. Lapo Cheli

Dean of the Doctoral Programme

Prof. Ing. Giovanni Ferrara

Supervisors

Prof. Ing. Carlo Carcasci

External referees

Prof. Ing. Gabriele Comodi

Ing. Paolo Colbertaldo

Years 2019/2022

Alla mia famiglia, a Chiara e a tutte le persone più vicine e care...

Ringraziamenti

In primo luogo, vorrei ringraziare il Prof. Carlo Carcasci, che mi ha accolto senza alcuna esitazione come suo dottorando, ha creduto in me, mi ha guidato in questo percorso difficile, ma stimolante, e mi ha dato l'occasione di poter crescere in questi anni, prima di entrare a far parte del mondo del lavoro e della ricerca.

Uno speciale ringraziamento va al Prof. Gabriele Comodi e all'Ing. Paolo Colbertaldo, per la loro disponibilità e il prezioso contributo nel miglioramento di questo lavoro di tesi tramite le loro revisioni e i loro consigli.

Un grazie di cuore ai miei compagni di dottorato, con cui ho condiviso fianco a fianco in questi anni avventure e disavventure: per aver affrontato insieme il lavoro durante la ripresa dalla pandemia, per il supporto reciproco nei momenti più duri e bui, ma anche per tutte le citazioni, le battute e i momenti indimenticabili.

Infine, vorrei ringraziare tutte le persone che mi hanno supportato lungo questo percorso: la mia famiglia, che ha sempre creduto in me e che mi ha trasmesso la passione e la perseveranza per superare i momenti difficili, Chiara, che è ogni giorno fonte inesauribile di conforto e di ispirazione con la sua dolcezza, ma anche con la sua determinazione e tutti i miei amici più cari, che ci sono sempre stati per me e che ci saranno sempre.

"Today I saved a life, my own. Am I a hero? I really can't say, but yes!" Cit. Michael Scott

Summary

Global warming is a severe problem for our planet. In the last decades, sustainable policies of countries, climate actions, and commission strategies have led to substituting hydrocarbon fuels, such as oil and coal, with Natural Gas (NG) which is a fossil fuel with the lowest carbon and environmental impact. However, this solution on its own will not be able to satisfy the targets defined by the EU 2050 long-term strategy. Europe and Italy in particular are large consumers and importers of NG and this means that their transport and distribution networks are extremely ramified and capillary. The Italian Natural Gas Distribution Network (NGDN) includes thousands of NG metering and pressure reduction stations, called City Gate Stations (CGS), for injecting gas into low-pressure networks from high-pressure transport networks. These plants are mainly based on the constant-enthalpy throttling of the gas flow to reduce its pressure, which leads to a significant reduction of its temperature by the Joule-Thomson effect. Gas preheating systems that avoid excessive cooling are installed upstream of pressure reduction valves and usually exploit conventional gas boilers. For this process, not negligible amounts of thermal energy are required, usually obtained by burning part of the gas flow rate. In addition to the necessary and rightful urgency of reducing the carbon footprint of the natural gas infrastructure, the objective of containing the NG consumption could also help the system to be more resilient to price variations which in recent months have reached unprecedented levels, hurting the European economy and exacerbating energy poverty conditions. This Ph.D. thesis work pursued two different approaches to address this challenge and to contribute to its resolution: the development of white-box models, i.e., based on the first principles of thermodynamics, to carry out techno-economic feasibility analyses and propose solutions to decarbonize preheating demand by lowering natural gas consumption, and, on the other hand, the development of black-box, or data-driven, models to replicate the behavior of the analyzed systems and thus build digital twins that can be useful for monitoring the energy consumption of the systems themselves, as well as for evaluating the impact of possible energy efficiency solutions. First, the work started with the selection of a cluster of real plant datasets belonging to a partner local DSO. Eventually, 7 CGS are selected, with maximum operative gas flow rates and pressure drop varying from 300 Smc/h to 6000 Smc/h and from about 23 barg to more than 60 barg, respectively. Results of the data analysis showed that the peculiarity of these plants is the strong dependence on seasonality, with a clear division between thermal summer and winter and, at the same time, a wide variability in size and plant layout, especially considering the gas preheating system. Furthermore, there was a strong correlation between the gas flow rate consumed for preheating and the plant's operating conditions, namely the gas inlet conditions, such as flow rate, pressure supplied by the Transmission System Operator (TSO), and gas temperature at the inlet point. An ad hoc thermodynamic model was developed to estimate the thermal energy demand for preheating, exploiting experimental data from various real plants and simplified models of heat pumps and renewable systems. The super-simplified models already present in the literature were expanded with an accurate calculation model of the Joule-Thomson coefficient, based on the NIST regulation, a model based on the data provided by the DSO to estimate the gas inlet temperature, and the knowledge of the partner DSO

and the author on real operating conditions of such systems. The model was not suitable for detailed consumption estimation of an individual system, but it has proven to be a sufficient tool for estimating preheating consumption from a techno-economic feasibility analysis perspective. This model was flanked by: PV output calculation tools, air-water heat pumps and gas boilers models, and economic investment estimation tools, including the energy-saving certificates calculation. Three different plant layouts were analyzed to decarbonize the system: off-grid, partially on-grid with only sale opportunity, and on-grid with the possibility of the grid purchasing to provide the heat pumps with electricity. Once the optimal techno-economic solutions were found, they were subjected to sensitivity analyses to changes in electricity and gas prices. Results showed that it is possible to install heat decarbonization systems based on a combination of PV and heat pumps with low payback times, starting from 3 to 5 years. However, the solutions are very dependent on the fluctuation of absolute prices and the ratio between the price of gas and that of electricity and it has not been possible to decarbonize 100% of the pre-heating without distancing from the optimum of the techno-economic analysis. Considering the problem from another point of view, it is not enough to propose viable plant solutions, but also to monitor whether their implementation impacted the decarbonization of the monitored plant. In addition, monitoring preheating also allows the system to be kept safe from failures, as insufficient preheating would lead to component damage and disruption to users downstream of the CGS. Finally, the difficulty of modeling the system accurately with 0-D models based on the first principles of thermodynamics opens the way to the possibility of solving this problem by developing data-driven models for each plant. Several regression-based machine learning models were developed and trained, a Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) algorithms. The selection of the inputs for these models was obtained through the features selection and engineering process. The algorithms were trained on a training data set relating to a complete year of operation for two real plants and their predictive performances were tested on another data set relating to the subsequent operating conditions of the same plants. The algorithms performed greatly, in terms of all the metrics chosen for their evaluation on the testing dataset. The algorithms, and in particular the ANN models, which are capable of understanding the non-linear interactions of the system, performed extremely satisfactorily on both datasets in terms of all the metrics chosen for their evaluation, such as R^2, RMSE. Values higher than 0.98 were obtained for the R^2 coefficient, with mean square percentage errors lower than 3% for both models and plants. The models were integrated into an energy monitoring system using the CUSUM technique and it was possible through these to identify malfunctions and waste within the testing period. This method allowed for the creation of "baseline" consumption models and to evaluation of the performance of the CGS over time using the CUSUM technique to find the variations between the actual and the modeled gas consumption, being an essential tool for monitoring the effectiveness of the natural gas preheating system. With this thesis work, methods and tools have been developed for the techno-economic analysis of efficiency and decarbonization solutions and machine learning-based monitoring of CGS for NG reduction. This will allow support and help DSOs, providing them with a series of tools that can be useful to address, acting immediately, the challenge of decarbonizing gas infrastructures.

Nomenclature

Acronyms

ANN Artificial Neural Network

$ASHP$ Air Source Heat Pump

BU Boiler Unit

$CAPEX$ Capital Expenditure

CD Coefficient Of Determination

CF Cash Flow

CGS City Gate Station

CHP Combined Heat and Power

COP Coefficient Of Performance

$CUSUM$ Cumulative Sum

DSO Distribution System Operator

EEC Energy Efficiency Certificates

FM Fiscal Measurment

LHV Lower Heating Value

MAE Mean Absolute Error

$MAPE$ Mean Absolute Percentage Error

ML Machine Learning

MLP Multi-Layer Perceptron

MLR Multiple Linear Regression

MSE Mean Squared Error

NG Natural Gas

$NGDN$ Natural Gas Distribution Network

NN Neural Network

NPV Net Present Value

OD Odorizer

$OPEX$ Operational Expenditure

PBT Pay Back Time

PH PreHeater

PI Profit Index

PRS Pressure Reduction Station

PV Photovoltaic

$RMSE$ Root Mean Squared Error

SCR Self Consumption Ratio

SSR Self Sufficiency Ratio

TSO Transmission System Operator

Latin symbols

C Specific cost, $€/kW$

c Specific heat, kJ/kgK

E Energy, kWh, kJ

G Irradiance level, W/m^2

I Investment, $€$

P Pressure, bar_g

Q Volumetric flow rate Smc/h

r Correlation coefficient, kJ/kgK

r Interest rate , $\%$

S CuSuM, -

T Temperature, $°C$

V Volume, m^3, L

W Power, kW

Greek symbols

η Efficiency, $(-)$

μ Joule-Thomson coefficient, $°C/MPa$

ρ Gas density, kg/Smc

Subscripts and superscripts

el electrical

gas	natural gas
$grid$	electricity from grid
H	constant enthalpy
hp	heat pump
in	inlet
JT	joule-Thomson
$lift$	temperature lift
max	maximum
min	minutes
out	outlet
p	constant pressure
$peak$	power peak
ph	re-heating
ref	reference
$sell$	electricity to grid
$soil$	buried pipeline surrounding soil
SP	set point
st	standard conditions
th	thermal
wat	preheating water
y	yearly

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Global warming is a severe problem for the planet. In the last years, sustainable policies of countries, climate actions and commissions strategies have led to substitute hydrocarbon fuel, such as oil and coal, with natural gas which is lower carbon fuel. While burning NG does emit greenhouse gases, it contributes far less CO2 and air pollutants than many of the fuels it is increasingly replacing, especially coal. Natural gas now accounts for about a quarter of global electricity generation. While in the medium term it is seen playing a major role supporting a transition to net zero energy systems, its longer-term use is uncertain in a world dominated by non-emitting renewable energies. According to the last IEA reports of [1], in the Net Zero Emissions by 2050 Scenario, unabated gas-fired generation continues its growth in the short term, displacing coal-fired generation, but starts falling by 2030 and is 90 % lower by 2040 compared to 2020. Increasingly, existing gas-fired power plants will need to be retrofitted with Carbon Capture systems or co-fired with low-carbon fuels such as hydrogen to be consistent with Net Zero Scenario levels. The depth and intensity of today's crisis have led to concerns about the future cost and availability of NG which have damaged confidence in its reliability and put a major dent in the idea of it serving as a transition fuel. As a result, the era of rapid global growth in NG demand is drawing to a close. In the Stated Policies Scenario (STEPS), demand rises by less than 5% between 2021 and 2030, compared with a 20% rise between 2011 and 2020. It then remains flat from 2030 at around 4 400 billion cubic metres (bcm) through to 2050, with growth in emerging market and developing economies offset by declines in advanced economies. According to the above-mentioned report, in the Announced Pledges Scenario (APS), demand soon peaks and is 10% lower than 2021 levels by 2030. In the Net Zero Emissions by 2050 (NZE) Scenario, demand falls by 20% to 2030, and is 75% lower than today by 2050. Theese three scenarios for the future of Natural Gas usage are showed in 1.1.

The projections released by the IEA 2022 report suggest that the peak of NG consumption in the world is yet to come, and it will probably be around 2030-2033; after this peak, the NG world consumption will face a constant reduction, but this fossil fuel will still be present in 2050. Since the progressive substitution of fossil fuel energy sources with renewable ones will require decades, eliminating unnecessary energy consumption could reduce CO_2 emissions in the medium term. Natural gas (NG) will be the last fossil fuel to be abandoned because of its low carbon content and wide availability. Thus, it is worthwhile to pay particular attention to reducing NG consumption, especially in the distribution network. This chapter will show the Italian and European context of NG, with a particular focus on gas transport and distribution networks. Subsequently, a component of this infrastructure will be analyzed in detail and its importance and relative interest on the part of the scientific community will be explored in greater depth. Finally, the objective and methodology of the thesis will be presented.

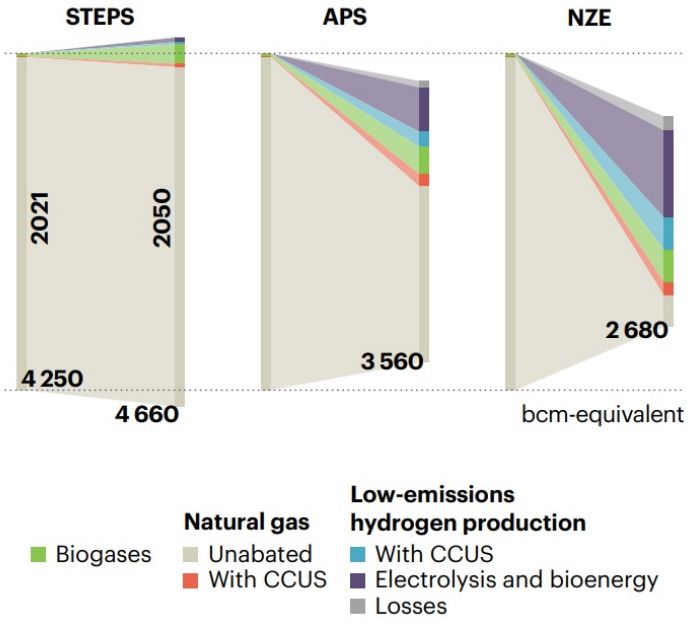


Figure 1.1: Comparison of the three world natural gas consumption scenarios according to [1]

1.1 Natural gas: European and Italian scenario

In the European Union in 2022 the NG baseline consumption was of about 360 bcm (billions of cubic meters). On the other hand. gas prices in Europe have reached unprecedented levels in the last year. That is hurting the European economy, compromising its ability to completely overcome all of the dramatic issues associated to the post-pandemic era. Moreover, the continuous pay rise in energy prices exacerbates energy poverty conditions by increasing the number of people in need and widening inequalities. Figure 1.2 shows how this baseline consumption is divided: about half is imported through pipelines, a little less than half through gas liquefaction (LNG) systems, and the remainder is obtained from domestic production.

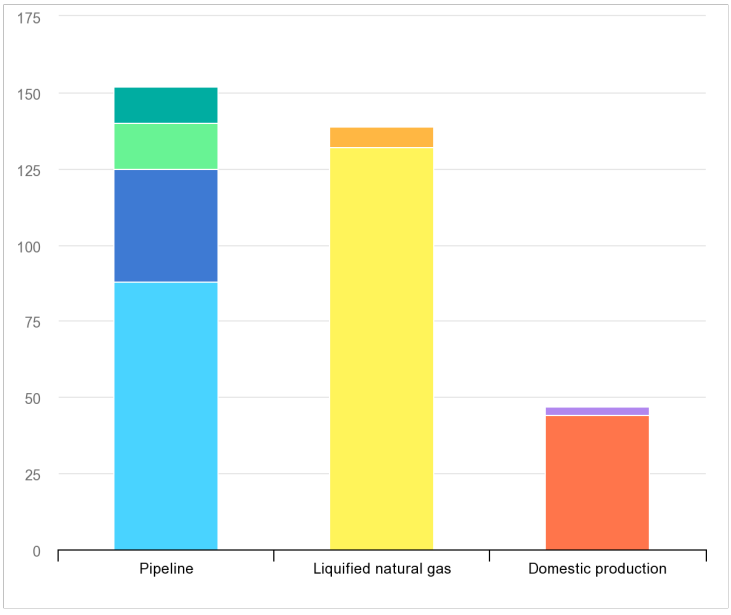


Figure 1.2: Baseline level of gas supply available to the European Union in 2023 [1]

Focusing on the national scenario, Italy is a huge importer of NG form different origins, such as Lybia,

Tunisia, Northen Europe, etc...According to Figure 1.3, Natural Gas in 2021 accounted for more than 40% of the share of primary energy supply. One of the main reason why Italy is so dependent on Natural Gas is because is the first user of this fossil fuel for generating electricity (see Figure 1.4) and the heating sector is still mainly based in buring NG in convetional boilers. As a consequence of this fact, the gas transport and distribution network in Italy is extremely important and widespread.

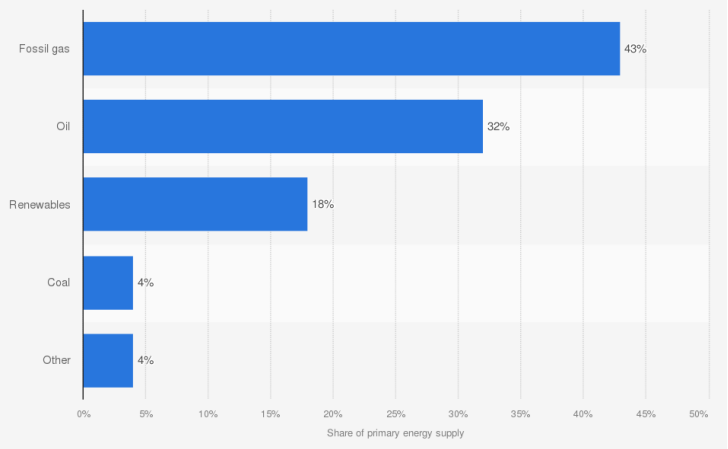


Figure 1.3: Distribution of the total primary energy supply by source in Italy in 2021 [2]

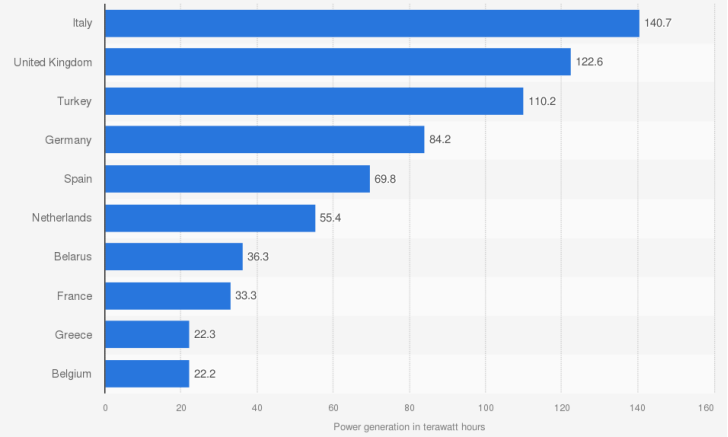


Figure 1.4: Leading country by gas electricity generation in Europe in 2021 [2]

1.2 Italian natural gas infrastructure

The Italian natural gas distribution system is composed of several components: pipelines, compression stations, City Gate Stations (CGS), Pressure Reduction Stations (PRS), control valves, etc... The Italian gas network accounts more than 32,800 km of pipelines at various pressure levels, more than 13 recompression stations with a total installed power of more than 900 MW and more than 9000 monitored CGS of a multitude of different sizes and layouts. Currently, the volume of gas transported annually by the Italian transport network is approximately 75.40 billion cubic metres [3]. Figure 1.5 is obtained directly from the Italian TSO website and shows which are the main import points and routes of major inter-regional and regional pipelines across the Italian peninsula.

The entire network is divided into two main parts:

- The national High-pressure network: including the systems involved in the transportation of natural gas from the injection points to the regional interconnections and storage sites. It consists of

Figure 1.5: Italian transport grid map [3]

pipes, usually of large diameter, with the function of transferring quantities of gas from the entry points of the system (imports and main national productions) to the points of interconnection with the regional transport network and with the storage structures. Some inter-regional gas pipelines functional to reaching important consumption areas are also part of it. The transport operator in italy, or Transport System Operator (TSO) is SNAM.

- The regional Low-pressure network, including the systems required for the local transportation of natural gas and the supply of industrial/urban users and power plants. The Distribution System Operator (DSO), on the other hand, is a figure that changes by region and sometimes even by

province, depending on what areas of the territory are covered.

Pipes are the elements of the network through which the gas is transported and distributed. Gas networks and consequently pipes are classified [4] according to the working pressure of the gas, as shown in table 1.1. Gas network pipes can also be divided up into high-pressure (1a, 2a, 3a species), medium-pressure (4a, 5a, 6a species) and low-pressure (7a species) pipes. High-pressure pipes are pipelines used for long-range gas transportation. Medium/low-pressure pipes are responsible for distributing gas in urban zones.

Table 1.1: Pipes species and relative working pressures according to Italian UNI.

	Transportation				Distribution		
Species	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}
P bar_g	>24	24 ÷ 12	12 ÷ 5	5 ÷ 1.5	1.5 ÷ 0.5	0.5 ÷ 0.04	<0.04

Figures 1.6 shows a real transportation and distribution pipes of a gas network. Pipelines are tubes of large diameter (up to 1.200 m) made from unprotected iron and carbon steel. Medium/low-pressure ducts are manufactured by high/medium-density polyethene or polytetrafluoroethylene lined carbon steel or carbon steel (old tubes). Distribution pipe diameter depends on gas pressure and flow rate. Typically, values are included between 80 mm and 300 mm.

(a) (b)

Figure 1.6: Transport pipeline (a) and local distribution pipes (b) examples.

Gas standards determine a maximum gas flow velocity for transmission and distribution pipes' species. Velocities into the pipes of the network should be lower than maximum allowed values to minimize pressure drop, impurity dragging and noise phenomena. The overcoming of the velocity limit can produce undesirable high-pressure drops in the network and consequently inadmissible pressure values at demand nodes.

Figure 1.7 shows all the components of the Italian Natural Gas infrastructure. The element embedded in red is the one that couples these two infrastructural levels and it is called in Italian the "ReMi" (Regolazione e Misura in Italian, which means Regulation and Measurement) cabin, or City Gate Station (CGS) in the literature, which means that facility equipped with a series of components to regulate the pressure of natural gas as its main task. Moreover, the CGS absolves also to several other purposes, such as natural gas odorization, which is a key task to ensure the detection of gas leaks, fiscal measurement to evaluate the amount of energy delivered to the users located downstream of the CGS, gas filtering to avoid impurities to enter the pressure reduction stage and gas preheating. The CGS is the physical connection

point between the national transport network and the local distribution network. The gas is taken from the national and inter-regional transport networks at high pressure whose values vary approximately from 10 to 70 bar. At these pressure values, the gas is not suitable for distribution and so it is necessary to carry out an initial pressure reduction in the "ReMi", also known as first-stage cabins. Then it represents:

- A delivery point for the transport network
- A withdrawal point and the gas source for the distribution network

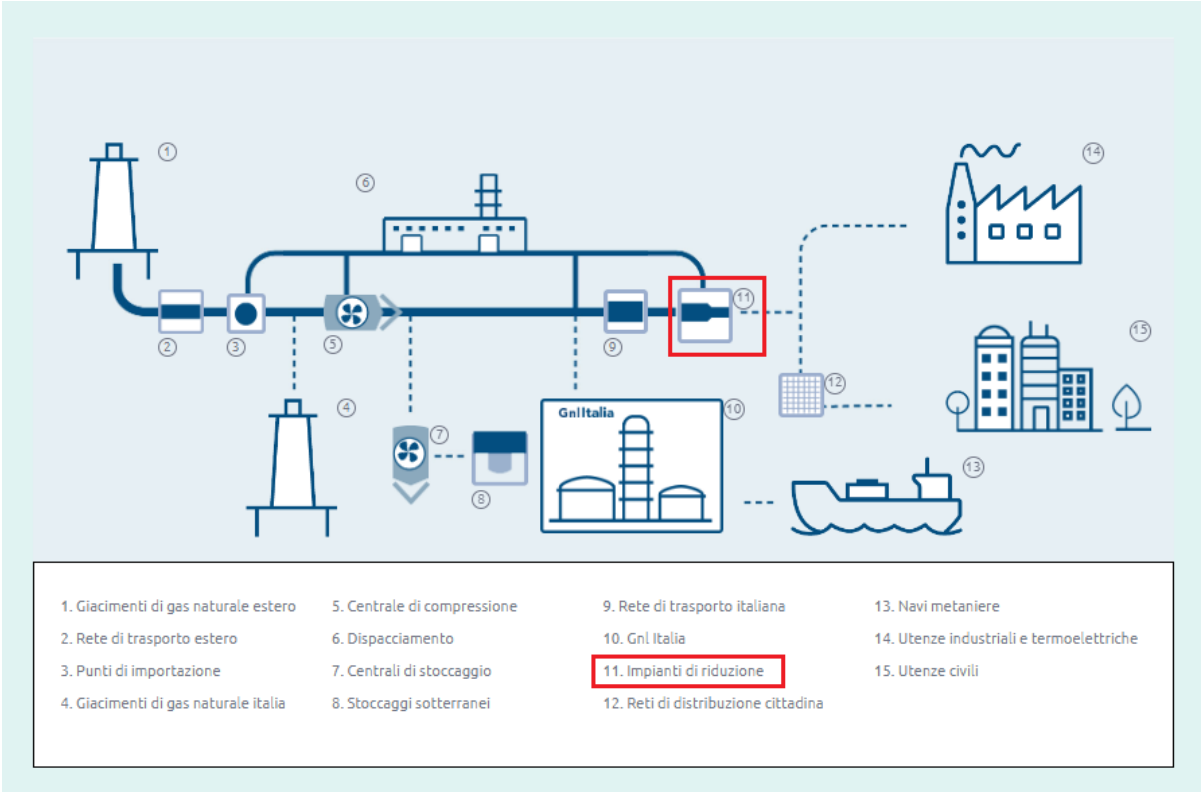
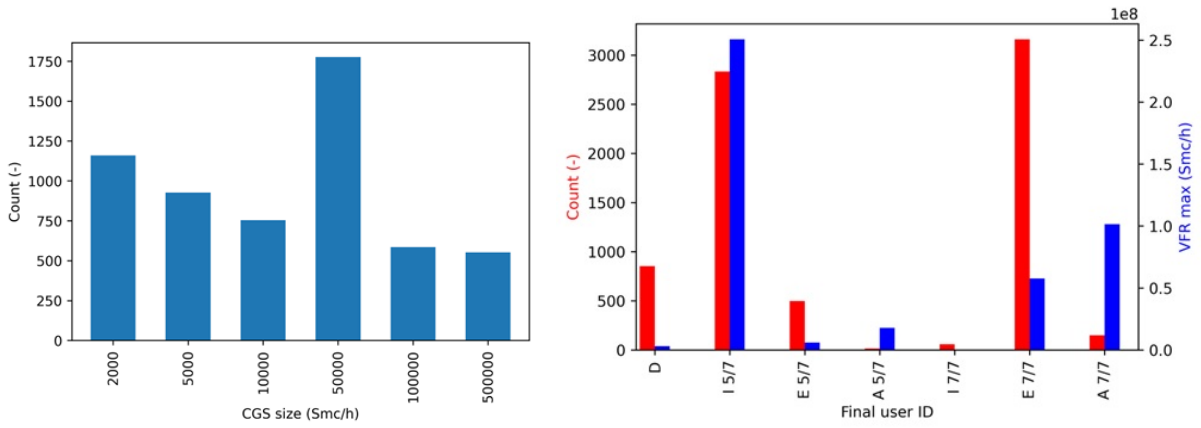


Figure 1.7: CGS location inside NG infrastructure [3]

It is very hard to obtain the hourly consumption profiles of all the plants present on Italian territory, however Snam [3] provides a list of all the CGS present in Italy correlated with the values of allocated flow rate (which can be understood as the maximum flow rate that can be supplied by the TSO and that can cross the transmission network), pressure drops and so on. It is interesting to note from figure 1.8 that most plants in Italy are small to medium-sized, i.e. with maximum rated flow rates less than or equal to 50000 Smc/h, which then correspond in practice to much lower maximum annual flow rates (even half), as the allocated flow rate differs greatly from the actual flow rate.

There are different types of connected consumers to the CGS: industrial (I), distribution (D), autotraction (A), thermoelectric (E) and so on[5], below what the authority classifies as consumption profiles:

- D': 'Distribution network'
- I 5/7: Industrial and other uses with offtake days equal to 5 out of 7.
- E 5/7: Thermoelectric with offtake days equal to 5 out of 7.
- A 5/7: Autotraction with offtake days equal to 5 out of 7.
- I 7/7: Industrial and Other with offtake days equal to 7 out of 7.



(a) (b)

Figure 1.8: Italian TSO - Delivery points analysis: Segmentation of the flow rates processed by the various plants and analysis of the types of users connected downstream of these plants.

- E 7/7: Thermoelectric with offtake days equal to 7 out of 7.
- A 7/7: Autotraction with offtake days equal to 7 out of 7.

1.2.1 City Gate Station layout

As mentioned earlier, the CGS performs many different tasks within the NG infrastructure.

1. Upstream inlet step: including the section of pipe ranging from the point of delivery to upstream manifold of the filters. They are also included: shut-off valve, insulating joint, emergency valve. The gas velocity must be less then 30 m/s.
2. Filtering step: intended for the separation of liquid and/or solid particles possibly present in the gas. The load losses of the clean filter must be less than 3 % of minimum inlet pressure
3. Preheating step: the natural gas preheating process it is required by law for powered systems from an inlet pressure of 12 bar. It must prevent excessive lowering of temperature produced by the relevant jump of reduction of gas pressure The gas temperature must not be lower than 5°C according to safety regulations. The formation of ice would cause serious problems operation of the regulating devices until reaching even when blocking the supply. Hot water generally produced with gas burners that exploit part of the gas is used as fuel processed
4. Pressure regulation step: ts purpose is to reduce the pressure of the gas in transit to a value constant. It is the main element taht includes, in addition to service regulators, also the control and emergency regulators, any locking devices, vent valves to the atmosphere, related accessories, pilots and connections, pressure ports, relief and safety valves and so on.
5. Fiscal measurment step: Measurement of output quantities (pressure, volumetric flow rate, temperature).This step guarantees fiscal measurement of gas in transit in the RE MI cabin. Are allowed :deformable wall meters, mostly for private users or in any case small flow rates, turbine meters, rotary counters, ultrasonic counters, diaphragms with pressure difference meters, Venturi pipes, etc...
6. Odorization: Natural gas is odourless and thus an odor is required for leak detection. Substances called odorants are used. Two kind of possible technologies: lapping odorization, injection odorization; usually theese two are used in combnation and the lapping acts as a reservoir for injection.

Figure 1.9: CGS filtering stage.

7. Downstream outlet step: it is the section that connects the exit of the group measurement with the shut-off valve general output of the plant. Connect the CGS to the network of distribution. Output pressures typically at values of pressure between 12-10 and 1-5 bar. It comprehends the downstream emergency valve and the shut-off valves and the insulating joint.

Figure 1.13 shows the trend of natural gas conditions in terms of pressure and temperature inside the CGS and along the path followed by the gas itself. The gas generally enters between 12 and 75 barg or between 1.2 MPa and 7. 5 MPa with a temperature varying from the area and topology of the upstream

Figure 1.10: CGS preheating heat exchangers and monitors for pressure reduction stage.

Figure 1.11: CGS fiscal measurment stage.

section, then it is filtered and its condition remains almost unchanged except for a small pressure drop after filtering, then it is preheated and the temperature rises to 40-60°C, then it undergoes the actual lamination and both P and T decrease sharply, and finally it passes through the various metering and odorization stations and exits in useful condition to be transported downstream of the CGS by the local distribution network.

1.2.2 Natural gas withdrawal profiles and climate zones

In Italy, there are different climate zones depending on the region and geographical area; depending on whether the CGS is located in one area rather than another, there will be different thermal seasons (winter,

Figure 1.12: CGS odorization stage: injection and lapping systems in series.

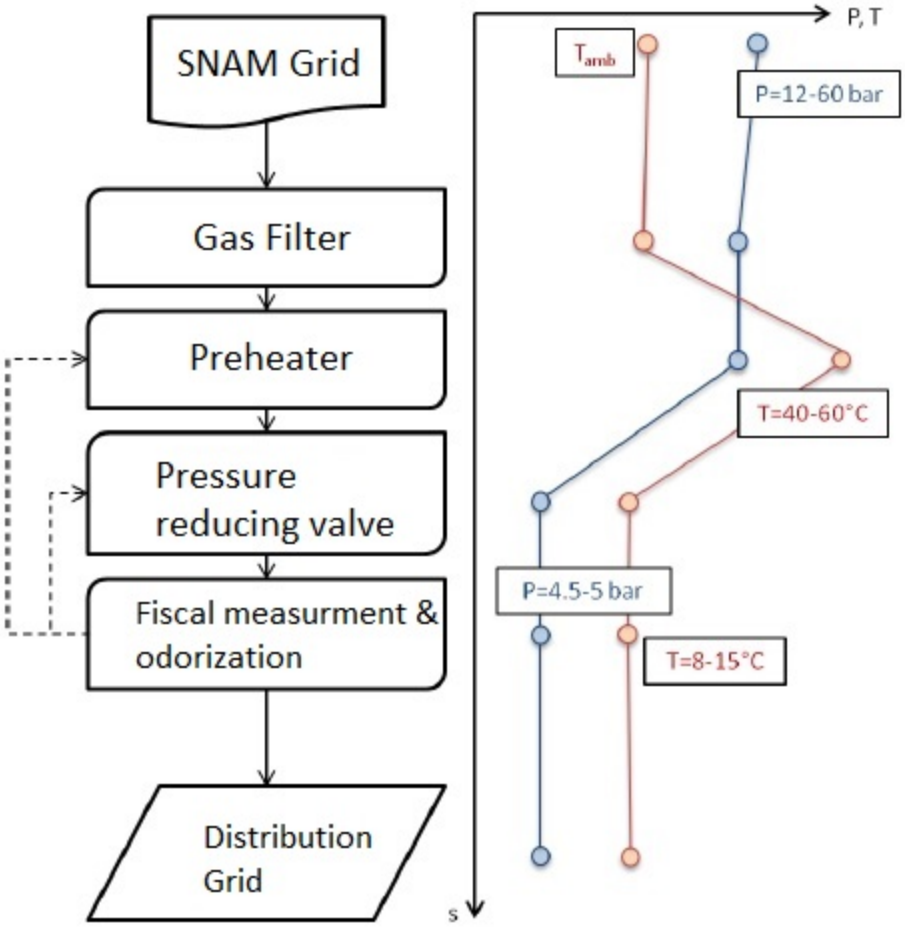


Figure 1.13: Italian CGS typical layout steps.

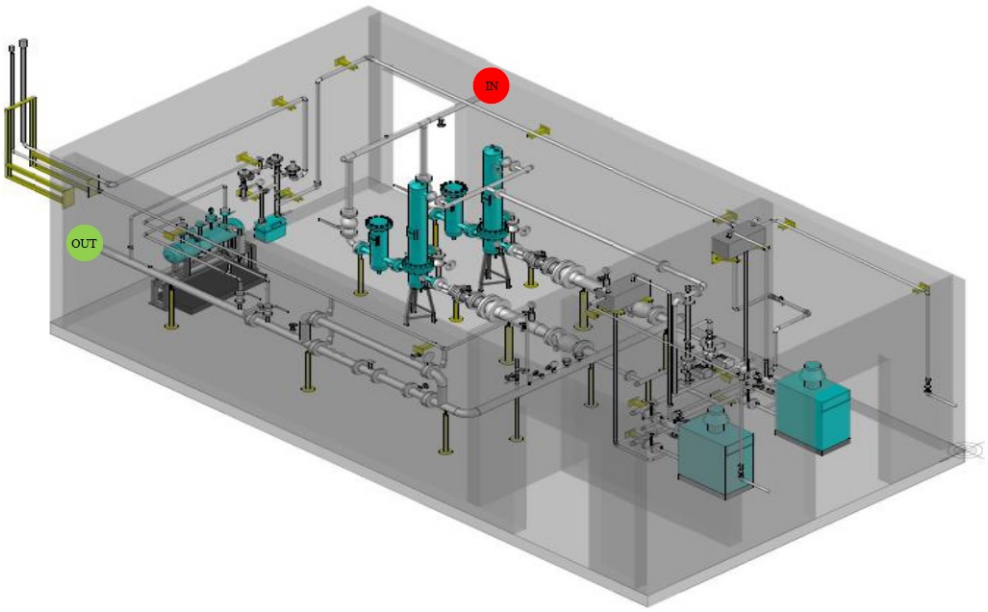


Figure 1.14: Italian CGS typical layout: real example.

where heating is allowed on, and summer, where heating is turned off) of different durations. Figures 1.15 shows the different areas in Italy corresponding to different climate zones.

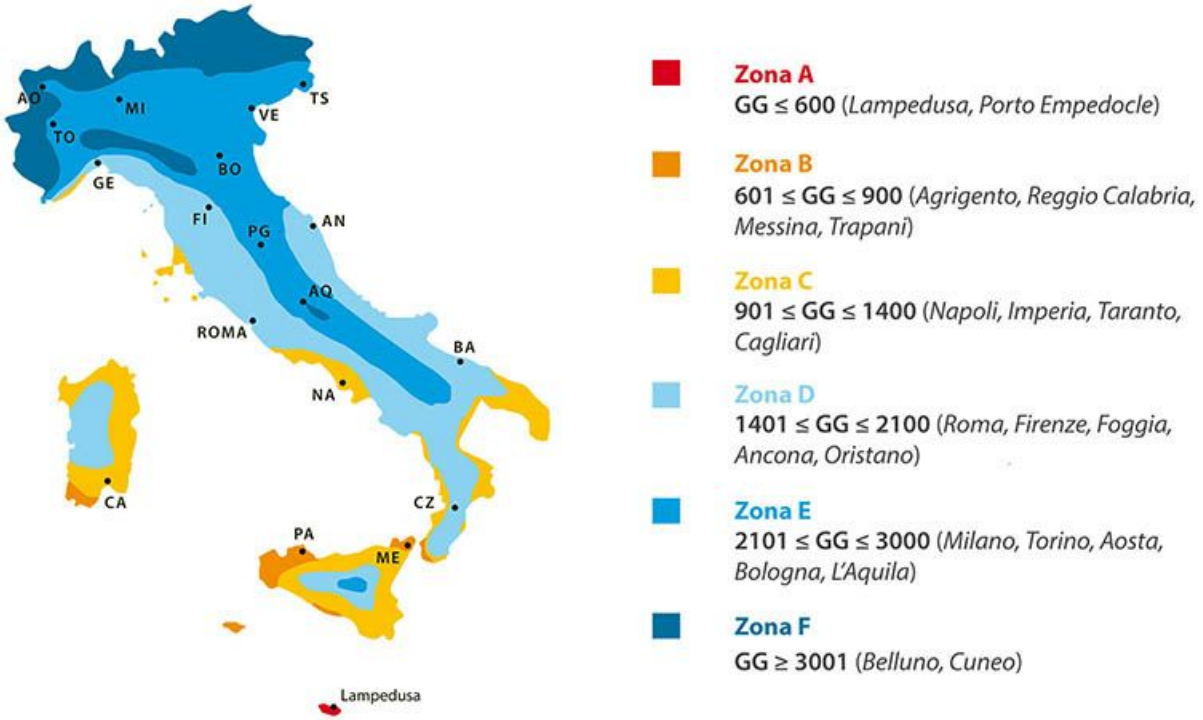


Figure 1.15: Climate zones in Italian area according to heating degree days.

The 7 CGSs used for this analysis all belong to the most common climate zones in Italy, i.e. zones D and E. The next figure (1.16), on the other hand, shows the periods when the heating is switched on and off for each climate zone. This factor is important and has an impact on the operating conditions of the CGS, especially when the users downstream of the CGS are mostly residential/commercial or in any case not industrial only and therefore strongly influenced by seasonality and the presence or absence of gas heating.

Climate regions are based on the concept of degree days, described by the 1.1 equation.

CLIMATIC ZONE	ON-OFF PERIOD	SCHEDULE
ZONA F (Trento e zone alpine)	**Unlimited**	**Unlimited**
ZONA E (Milano, Torino, ecc.)	**22 October – 07 April**	**13 hours a day**
ZONA D (Roma, Pescara, ecc.)	**08 November – 07 April**	**11 hours a day**
ZONA C (Napoli, Caserta, ecc.)	**22 November – 23 March**	**9 hours a day**
ZONA B (Palermo, Trapani, ecc.)	**08 December – 23 March**	**7 hours a day**
ZONA A (Lampedusa, Sud e Isole)	**08 December – 07 March**	**5 hours a day**

Figure 1.16: Times and periods of switching on heating in italy according to the network regulation authority [5]

$$HDD = \sum_{i}^{n} (T_0 - T_{env}) \tag{1.1}$$

Where T_0 is the indoor environmental temperature, T_{env} is the external temperature, i is the pedix for the considered day and n is the number of days in one year, i.e. 365. Heating degree days (HDD) are defined by an Italian legislative decree as the "sum, extended to all days of a conventional annual heating period, of only the daily positive differences between the room temperature, conventionally set at 20°C, and the daily average outdoor temperature; the unit of measurement used is the degree day (DD)." In simpler words, degree days are the difference between the average daily temperature and the 20°C we have in the house, for all days of the heating period.

Thus, it is clear that the working profile of a CGS and consequently its preheating consumption profile will be influenced by some of the factors and conditions described in the previous sections: type of utilities, type of climate zone, and distribution of the type of utilities in the cluster of utilities downstream of the CGS.

1.2.3 Preheating requirements: Joule–Thomson Effect

The temperature change undergone by the gas during an adiabatic expansion depends not only on the final and initial pressure states, but also on the way the expansion is carried out. In a free expansion the gas does no work and absorbs no heat, so the internal energy is conserved. Expanding in this way, the temperature of an ideal gas would remain constant, but the temperature of a real gas decreases, except at very high temperatures. The Joule-Thomson expansion method, on the other hand, is an intrinsically irreversible method. During this expansion, the enthalpy remains unchanged, but unlike a free expansion, work is done that causes a variation in internal energy. This change due to the irreversibility of the process means that much greater cooling or heating can be achieved than in the case of free expansion. The

operating principle of the NG grid is based on sequentially controlled multiple gas expansions, commonly performed by throttling valves that dissipate the energy previously spent on NG compression. This dissipation of energy causes by the Joule-Thomson effect a reduction in the gas temperature as well, and this can cause plant engineering as well as technical problems for the entire network downstream of the CGS. Expansions involve cooling the gas and possible hydrate formation, potentially compromising the integrity of components installed downstream of the pressure reduction stations.This effect is compensated for by pre-heating systems that use part of the natural gas flow to feed conventional gas boilers that heat water that exchanges heat with the gas entering the CGS.

Figure 1.17 shows the P-H (Pressure - Entalphy) plane of methane, which is used to better understand this phenomenon. Although NG is a mixture of hydrocarbons, it is mainly composed of methane, the molecule with formula CH4,. From the graph, two points (A represents the starting of the expansion and B the arrival) can be seen highlighted; it can be seen that one moves in the area outside the bell in the gaseous state vertically (iso-enthalpy) and that a reduction in pressure consequently leads to a shift to another iso-temperature of lower value.

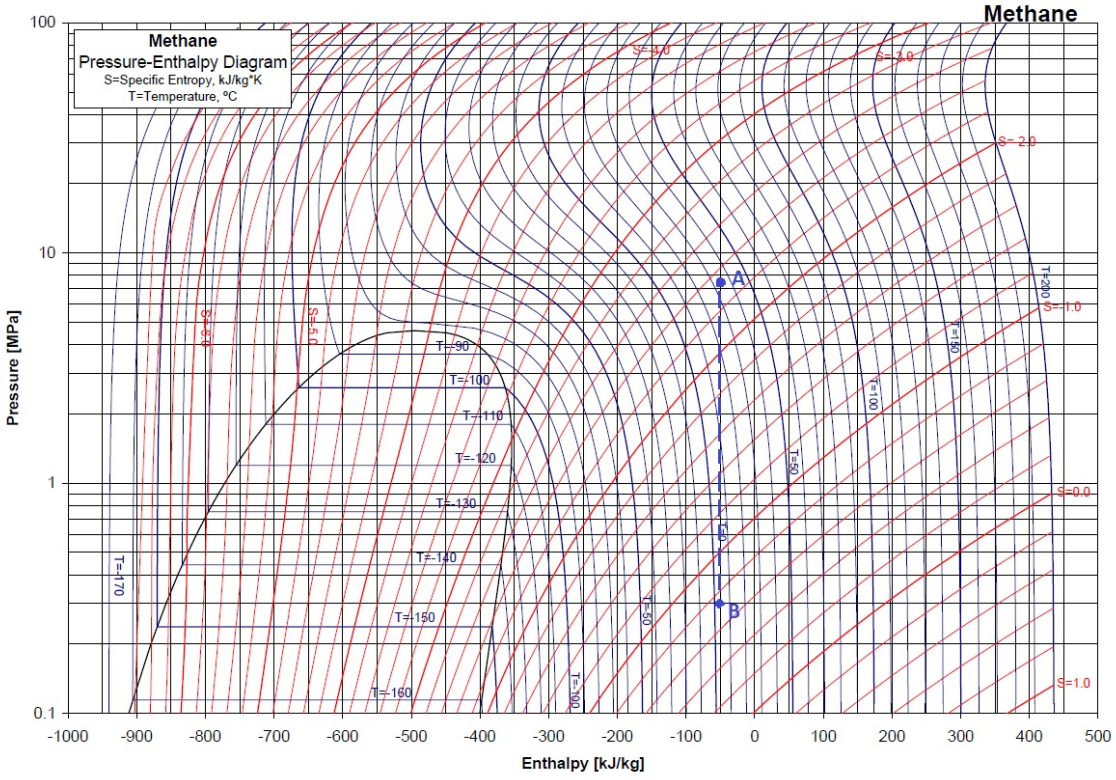


Figure 1.17: P-H diagram for methane gas

While before this energy dissipation, gas consumption and CO_2 emissions be considered acceptable and part of the "operating costs" of the gas infrastructure, nowadays with decarbonization targets becoming more and more stringent and rigid on the one hand, and the gas prices skyrocketing, on the other hand, it is crucial to act on these systems to reduce and if possible completely decarbonize their internal processes.

1.3 Literature review

The EU's climate neutrality goals include energy efficiency in the energy sector, and the NG sector, despite being based on the cleanest fossil fuel and having the longest survival horizon, must still ensure an increase in energy efficiency and progressive decarbonization. There are several areas to act on for the

decarbonization of the gas transport infrastructure, to name a few:

- decarbonization of the gas carrier through injection of green gases such as bio-methane [6, 7] or hydrogen [8–10];
- reduction of gas fugitive emissions into the atmosphere [11];
- energy recovery or energy efficiency in pre-heating systems in CGS (see tables 1.3 and 1.2).

CGS modelling and monitoring and NG preheating efficiency and energy recovery have already been studied in several scientific papers. Academic studies generally deepen two main approaches: energy recovery in the gas through expanders that exploit the residual pressure drop to produce electricity or systems for reducing or abating the energy cost of preheating the gas. A third strand, which is much poorer in terms of published papers, concerns the analysis and monitoring of these facilities using data-driven techniques or reliability and safety approaches. The studies that are more widespread concern techno-economic feasibility analyses involving different types of technologies and with the ultimate aims of both energy recovery analysis and pre-consumption reduction analysis.

Figures 1.19 and 1.18 show the results of the literature review for the CGS analysis. According to scientific papers, there are several technologies that could be implemented inside the CGS; these devices could be use in place of the conventional gas boilers or in support of them. The literature on these topics has grown rapidly in recent years, with several dozen papers on CGS coming out between 2018 and 2021. This proves that the attention paid by the scientific world to these topics is higher than ever.

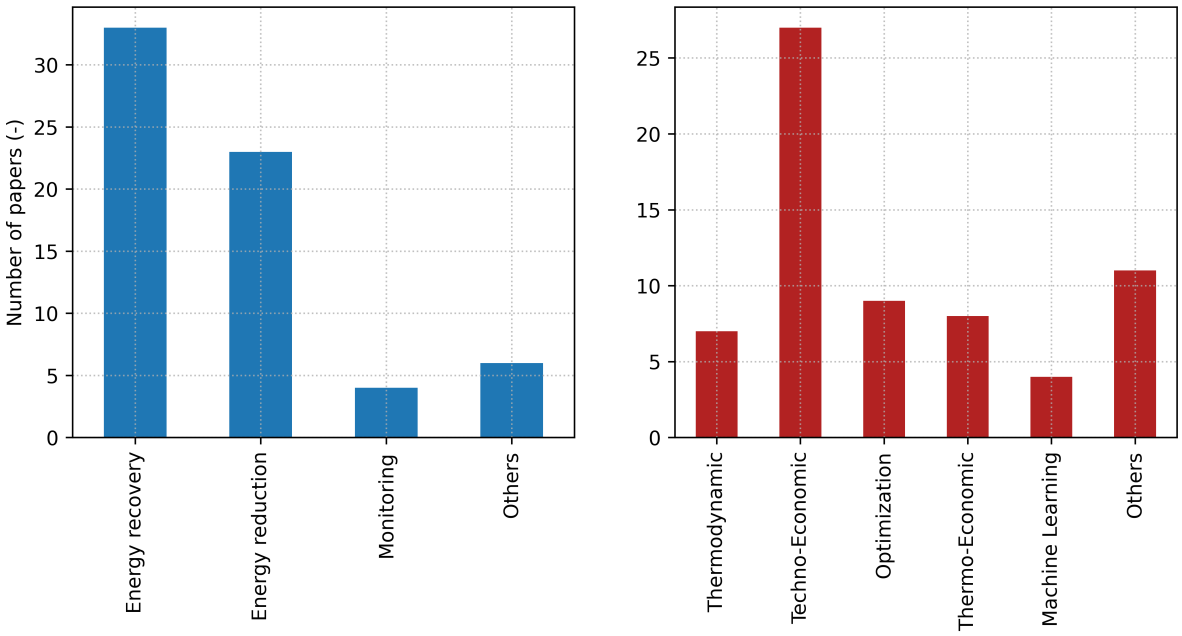


Figure 1.18: CGS literature review: analysis vs approaches.

One of the most common system upgrade consists of the installation of turbo-expanders to recover energy from the pressure cascades; generally speaking, these kind of devices allows a quick return of the investment if the CGS flow rates are very high and the pressure drops are average-low. Some plants are instead equipped with solar collectors and heat recovery systems to avoid burning fuel in the boiler for pre-heating needs. The high capital and operating costs can be justified in these cases if the demand curve (preheating gas consumption) and supply curve (solar radiation) are as coincident as possible. Considering the European and especially the Italian panorama, the majority of CGSs are of small to medium size. In

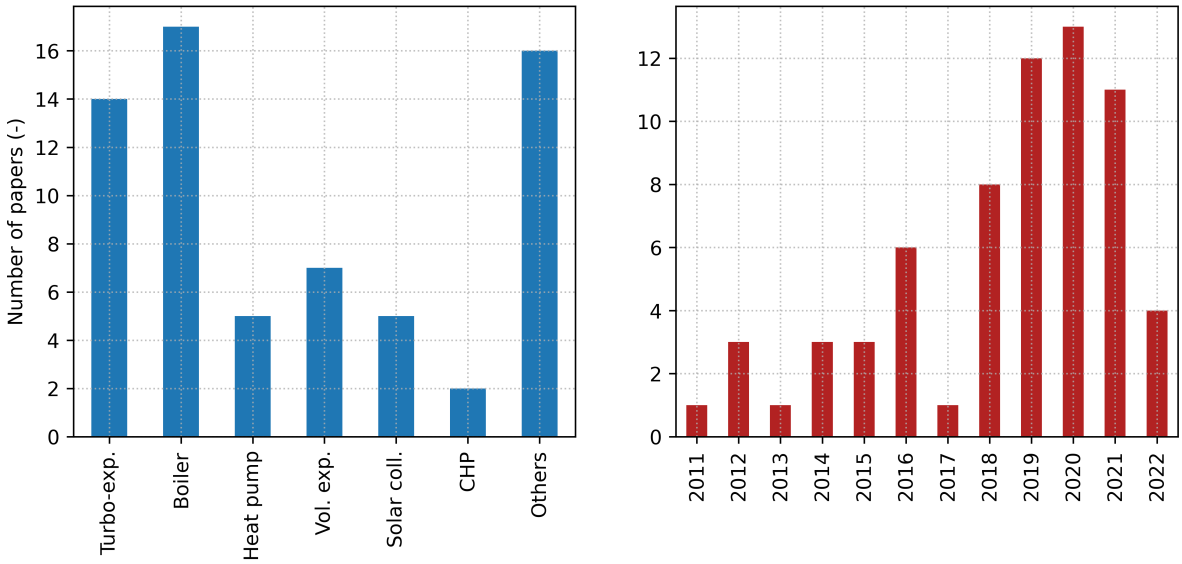


Figure 1.19: CGS literature review: techs vs years.

order to be able to recover energy from these plants, expansion systems based on volumetric expanders are designed, which are better suited to processing small flows and medium to large pressure drops. In this case, the reduced gain in terms of recovered energy is counterbalanced by a much lower capital cost compared to turbo-expanders. Heat pumps, air or geothermal, have been widely designed to support or completely replace gas boilers. The advantages of these technologies are many: two to three times the efficiency of boilers depending on the region of use and demand, lower capital costs than expanders, and a general technological trend that is continuing to lower prices given the wide spread in other industrial and civil sectors. The use of heat pumps to partially and fully decarbonise pre-heating costs is therefore an interesting concept that can still be explored further. Still in terms of technologies, there are, finally, some studies concerning more specific and particular technologies. To name a few: Ranque-Hilsch vortex tubes, Allam cycles, heat pipes, ORC cycles and so on.

1.3.1 Preheating reduction approach

Farzaneh-Gord et al. [12] propose a heat production system as a partial replacement for the traditional boiler consisting of a solar collector coupled to a tank, applied to a CGS placed in Akand. The authors find the optimal number of collectors and storage tank capacity based on the technical-economic analysis, in fact as the number of collectors increases, the fuel cost decreases, but the capital cost increases.

Farzaneh-Gord et.al. [13] then propose a new system to eliminate the fuel consumption of CGSs, using a ground-coupled vertical heat pump. The performance of the system is studied under two different climatic conditions in Iran and two different NG compositions. Results show that the system is fully capable of eliminating the preheating gas consumption by more than 65% and reducing CO2 emissions by up to 79%. The discounted payback period is computed to be around 2 years.

Deymi-Dashtebayaz et al. [14] present a new approach for optimizing fuel consumption in a CGS by defining the minimum possible controller output temperature. The minimum temperature is based on the gas hydrate temperature and was calculated by applying the fundamental thermodynamic equations and the equation of state, validating the model with the measured temperature of the CGS at Abbas Abad. Results show that the method can reduce gas consumption by approximately 35%, and the production of carbon dioxide due to incomplete combustion can be significantly reduced.

Englart et al. [15] propose the use of renewable energy sources in CGS Polish gas pre-heating to reduce

thermal energy consumption, analysing various combinations with conventional heat pump, absorption, and ground heat exchanger. Results highlight that the application of a gas heat pump to replace the traditional gas boiler could reduce gas consumption by up to 27-42% for the case study considered. The extension of the gas pre-heating system with an additional ground heat exchanger, used as a heat source for the heat pump, could lead to greater energy savings in gas consumption of between 30 and 44%. In a study for the following year, the authors focus on renewable energy source (RES)-based electrical technologies, such as air source and ground source heat pumps, coupled with air-to-ground heat exchangers, horizontal and vertical heat exchangers. The pre-heating estimation model is improved from previous work by considering the gas composition to estimate the basic properties of the fuel chemical compounds. Analyses were performed for three climate types (from cold to hot) and for the two operating modes. Results shows that the electric pre-heating solution with a RES system can save more than 50% of the primary energy, reducing greenhouse gas emissions [16].

Alizadeh et al. [17] study the possibility of improving the energy recovery efficiency in CGS using a heat pipe designed specifically for this purpose. This system is tested with real data from one year operation of pressure reduction stations, Results indicate that the heat pipes is capable to reduce gas consumption by more than half a million cubic meter a year and it annually prevents 756 tons of CO2 from being emitted .

1.3.2 Preheating energy recovery approach

Borelli et al. [18] investigate the integration of a CGS with low-temperature thermal energy sources by means of a two-stage expansion system. The risk of NG hydrate formation was evaluated for several Operating Conditions (OCs) with a transient model. The energy efficiency of the cabinets with low and high temperature configuration is compared. Results highlight that the expansion could achieve better energy performance and be integrated with low-enthalpy heat sources. The same authors investigate and propose key performance indicators for energy recovery in CGS, considering a theoretical reference process, in which a Joule-Thompson expansion and emission reduction indicators occur. Results showed that the proposed KPIs proved to be a useful, simple, and easily interpretable tool for managing the design development of heat recovery systems at CGS [19].

Danieli et. al. [20] study several kinds of mechanical expanders combined with different preheating devices based on gas boilers, cogeneration engines or heat pumps to identify the best combination by evaluating the combination of maximum net present value and minimum payback period, applied to Pressure Reduction Stations (PRS). Results show that small size volumetric expanders with low expansion ratios and coupled with gas-fired preheaters have the highest potential for large-scale deployment of energy recovery from PRSs with a maximum recover percentage of about 15% of the available thermal energy . In a following paper, the same authors evaluate the economic and technical feasibility analysis of a thermal energy recovery system based on the Ranque-Hilsch vortex tube. A model of the entire system is included in an optimisation method. A new empirical model of the device is proposed. Finally, a complete set of PRS from the Italian NG grid is chosen as case study, using the actual operating conditions collected by the DSO of each station. Results point out that the ambient temperature strongly influences the techno-economic feasibility of the proposed device, but 95% of pre-heating costs could be eliminated with a payback time of less than 20 years [21].

Mohammad Ebrahimi Saryazdi et al. [22] perform a multi-objective optimization of a NG preheating system composed of a turbo-expander supplied by a waste heat recovery device or a boiler unit. The total cost of the proposed configuration and the total exergy are used as objective functions. Results show that the configuration without the gas boiler unit is advantageous for both economic and exergy indicators.

1.3.3 Data-driven approaches

Few studies involve the use of data-driven approaches to these facilities; the main ones will now be described and are summarized in the table 1.4 and in figure 1.20. The use of data-driven approaches requires first the presence of a substantial data set, which is available to DSOs, but also a compelling need and a specific goal. Until now, there was no need to reduce gas grid consumption because of abundant supplies and low prices. The only data-driven approaches have involved plant health monitoring through fault detection or risk analysis approaches, with the sole objective of avoiding failures and malfunctions so as not to incur supply interruptions. With the rapidly changing global energy scenario, rising and fluctuating prices, and the demand for decarbonization, a need for energy monitoring aimed at controlling and reducing consumption by companies and DSOs has taken hold.

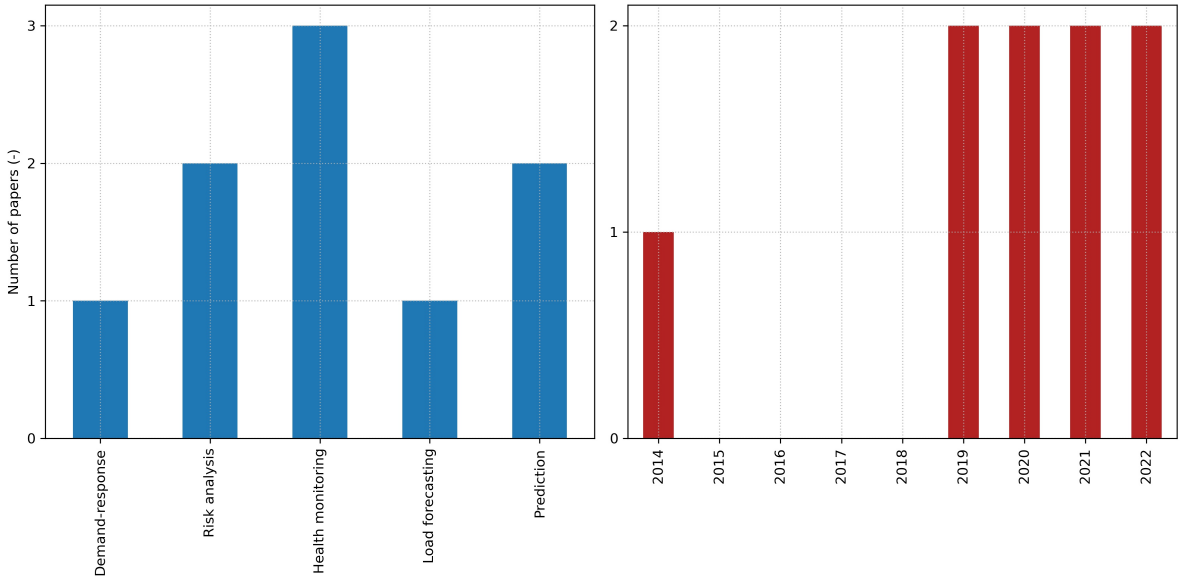


Figure 1.20: CGS data-driven methods literature review: approaches vs years.

Aramesh et. al. [23]present a novel approach to NG flow prediction at CGS using fuzzy-ANN by building a model based on data from a first station which is then subsequently tested to predict NG transmission from a second station and optimize the neural network error. The results showed that the predictive model was usable after adaptations even at a station that did not provide training data and the range of the transmitted NG volume was different. This approach was useful to develop data-driven predictive models of delivered NG in CGS with a lack of data but similar consumption to save building cost a new model for each CGS.

Farzaneh-Gord et.al. [24] propose a method based on Artificial Neural Networks (ANN), namely the Multiple Layer Perceptron (MLP), to predict the physical properties of NG as it transits through a CGS. Subsequently, the method is used to directly calculate the temperature drop along a throttling process. To train, validate and test the network, a large database of natural gas fields in Iran and some experimental data (30,000 random data sets) were collected from the literature. The results indicate that the developed machine learning methods exhibit high accuracy in calculations over a wide range of gas mixtures and input properties.

Bryan Leo et al. [25] developed a framework for uninterrupted condition monitoring of reduction and measurement cabins by developing a mathematical-physical model of the cabin behavior, which incorporates a real gas equation through a meta-model using an Artificial Neural Network (ANN). The results highlighted several potential health problems in a PRS and suggested various smart strategies to

control them by illustrating the application of the proposed method through a case study of a pressure reduction cabin with one year of data.

Sharma et al. [26] proposed a method for condition monitoring of a PRS, which was modelled using computationally cheap ML-based surrogate, avoiding the use of First Principles Models (FPM), which are computationally expensive. The gas filter chocking was monitored to predict its remaining useful life and a dashboard was implemented for health monitoring of the PRSs to be merged with online plant data for predictive maintenance aim. The results showed that this approach could increase the DS reliability and reduce maintenance costs.

Fan et al. [27] proposed a novel DR method for predictive management in multi-level natural gas markets with different stakeholders. This method included three parts: dynamic pricing model, intelligent decision making and data-driven demand forecasting. A Markov decision process-based model is developed to illustrate the process of dynamical optimizing energy prices. Then, deep learning and reinforcement learning are integrated to efficiently solve the sequential decision-making problem, based on the physics constraints of natural gas pipeline networks. The proposed dynamic pricing method was able to optimize the pricing strategies in accordance to the demand patterns, and dynamically improve the system stability and energy efficiency. Finally, they applyed the developed method to a natural gas network with relatively complex topology and different CUs. Results showed that the proposed method can achieve the targets of peak shaving and valley filling under different pricing periods.

Table 1.2: Energy recovery approaches for preheating

Title	Year	Techs	Reference
«An investigation of the performance of a hybrid turboexpander-fuel cell system for power recovery at natural gas pressure reduction stations»	2011	Turboexpander, Fuel Cell	[28]
«Producing electrical power in addition of heat in natural gas pressure drop stations by ICE»	2012	ICE	[29]
«Modeling and optimizing a CHP system for natural gas pressure reduction plant»	2012	CHP	[30]
«Recovery of Wasted Mechanical Energy from the Reduction of Natural Gas Pressure»	2014	Turboexpander	[31]
«Thermoecological cost of electricity production in the natural gas pressure reduction process»	2014	Expander	[32]
«Energy and exergy analysis of electricity generation from natural gas pressure reducing stations»	2015	Turboexpander	[33]
«Defining a technical criterion for economic justification of employing CHP technology in city gate stations»	2016	CHP	[34]
«Energy consumption minimization in an innovative hybrid power production station by employing PV and evacuated tube collector solar thermal systems»	2016	Solar Collector	[35]
«Performance assessment of vortex tube and vertical ground heat exchanger in reducing fuel consumption of conventional pressure drop stations»	2016	Vortex	[36]
«Proposal and design of a natural gas liquefaction process recovering the energy obtained from the pressure reducing stations of high-pressure pipelines»	2016	Other	[37]
«Feasibility study on energy recovery at Sari-Akand city gate station using turboexpander»	2016	Turboexpander	[38]
«Use of Rolling Piston Expanders for Energy Regeneration in Natural Gas Pressure Reduction Stations—Selected Thermodynamic Issues»	2017	Expander	[39]
«Energy recovery from natural gas pressure reduction stations: Integration with low temperature heat sources»	2018	Turboexpander	[18]
«Development and application of screw expander in natural gas pressure energy recovery at city gas station»	2018	Expander	[40]
«Key performance indicators for integrated natural gas pressure reduction stations with energy recovery»	2018	Turboexpander	[19]
«Optimal retrofitting of natural gas pressure reduction stations for energy recovery»	2018	Turboexpander	[41]
«Thermodynamic modeling and analysis of a novel heat recovery system in a natural gas city gate station»	2019	ORC	[42]
«Techno-Economic Assessment of Turboexpander Application at Natural Gas Regulation Stations»	2019	Turboexpander	[43]
«Optimal design and thermo-economic analysis of an integrated power generation system in natural gas pressure reduction stations»	2019	ORC	[44]
«Performance research on a power generation system using twin-screw expanders for energy recovery at natural gas pressure reduction stations under off-design conditions»	2019	Expander	[45]
«Thermodynamic and Economic Feasibility of Energy Recovery from Pressure Reduction Stations in Natural Gas Distribution Networks»	2020	Expander	[20]
«Comprehensive techno-economic and environmental sensitivity analysis and multi-objective optimization of a novel heat and power system for natural gas city gate stations»	2020	Other	[46]
«Optimal working conditions of various city gate stations for power and hydrogen production based on energy and eco-exergy analysis»	2020	Turboexpander, Fuel Cell	[47]
«Proposal and analysis of a coupled power generation system for natural gas pressure reduction stations»	2020	Expander	[48]
«Using the potential of energy losses in gas pressure reduction stations for producing power and fresh water»	2021	Turboexpander	[49]
«Optimal detailed design and performance assessment of natural gas pressure reduction stations system equipped with variable inlet guide vane radial turbo-expander for energy recovery»	2021	Turboexpander	[50]
«A Smart Energy Recovery System to Avoid Preheating in Gas Grid Pressure Reduction Stations»	2022	Vortex	[21]
«Energy, exergy, and eco-environment modeling of proton exchange membrane electrolyzer coupled with power cycles: Application in natural gas pressure reduction stations»	2021	Turboexpander	[51]
«Exergetic and economic evaluation of a novel integrated system for trigeneration of power, refrigeration and freshwater using energy recovery in natural gas pressure reduction stations»	2021	Turboexpander	[52]
«Energy, exergy and economic analysis of utilizing the supercritical CO(2)recompression Brayton cycle integrated with solar energy in natural gas city gate station»	2021	Brayton Cycle	[53]
«Thermodynamic analysis of a comprehensive energy utilization system for natural gas pressure reduction stations based on Allam cycle»	2022	Other	[54]
«Performance analysis of a power generation system for pressure energy recovery at natural gas city gate stations»	2022	Turboexpander	[55]
«Energy and exergy analysis and multi-objective optimization of using combined vortex tube-photovoltaic/thermal system in city gate stations»	2022	Vortex, PV	[56]

Table 1.3: Energy efficiency approaches for preheating

Title	Year	Techs	Reference
«Feasibility of accompanying uncontrolled linear heater with solar system in natural gas pressure drop stations»	2012	Solar Collector	[12]
«The minimum gas temperature at the inlet of regulators in natural gas pressure reduction stations (CGS) for energy saving in water bath heaters»	2014	Boiler	[57]
«Energy and exergy analysis of natural gas pressure reduction points equipped with solar heat and controllable heaters»	2014	Solar Collector	[58]
«Employing geothermal heat exchanger in natural gas pressure drop station in order to decrease fuel consumption»	2015	Heat Pump	[59]
«New energy-saving temperature controller for heater at natural gas gate station»	2015	Boiler	[60]
«Integration of vertical ground-coupled heat pump into a conventional natural gas pressure drop station: Energy, economic and CO2 emission assessment»	2016	Heat Pump	[13]
«Performance assessment of a novel natural gas pressure reduction station equipped with parabolic trough solar collectors»	2018	Solar Collector	[61]
«Energy and environmental analysis of a natural gas pressure reduction station equipped with turboexpander, solar collector, and storage tank»	2018	Turboexpander, Solar Collector	[62]
«Application of water reheating system for waste heat recovery in NG pressure reduction stations, with experimental verification»	2018	Other	[63]
«Optimization of fuel consumption in natural gas city gate station based on gas hydrate temperature (Case study: Abbas Abad station)»	2019	Boiler	[14]
«Renewable energy sources for gas preheating»	2019	Heat Pump	[15]
«Improving thermal performance of water bath heaters in natural gas pressure drop stations»	2019	Other	[64]
«Modelling of heat supply for natural gas pressure reduction station using geothermal energy»	2019	Heat Pump	[65]
«A novel modification on preheating process of natural gas in pressure reduction stations to improve energy consumption, exergy destruction and CO2 emission: Preheating based on real demand»	2019	Boiler	[66]
«Optimization of city gas network: a case study from Gujarat, India»	2019	Boiler	[67]
«Energetic, exergetic, environmental and economic assessment of a novel control system for indirect heaters in natural gas city gate stations»	2020	Boiler	[68]
«Reducing energy consumption for electrical gas preheating processes»	2020	Heat Pump	[16]
«Thermal modeling of indirect water heater in city gate station of natural gas to evaluate efficiency and fuel consumption»	2020	Boiler	[69]
«A comprehensive assessment of low-temperature preheating process in natural gas pressure reduction stations to better benefit from solar energy»	2020	Solar Collector	[70]
«Energy Harvesting by a Novel Substitution for Expansion Valves: Special Focus on City Gate Stations of High-Pressure Natural Gas Pipelines»	2020	Expander	[71]
«Performance enhancement of water bath heater at natural gas city gate station using twisted tubes»	2020	Boiler	[72]
«Investigation of Operational Scenarios to Mitigate CO2 Emission and Natural Gas Consumption in City Gate Stations (CGSs)»	2021	Turboexpander	[73]
«Energy analysis of vacuum tube collector system to supply the required heat gas pressure reduction station»	2021	Boiler	[74]
«An experimental investigation on using heat pipe heat exchanger to improve energy performance in gas city gate station»	2022	Other	[17]
«Analysis of the effects of forced convective heat transfer to reduce the efficiency of heaters of gas pressure reducing stations»	2022	Boiler	[75]

Table 1.4: Data-driven approaches

Title	Year	Approach	Reference
«A general neural and fuzzy-neural algorithm for natural gas flow prediction in city gate stations»	2014	Data-driven prediction	[23]
«Design and implementation of an expert system for periodic and emergency control under uncertainty: A case study of city gate stations»	2019	Health monitoring	[76]
«A fuzzy expert system for mitigation of risks and effective control of gas pressure reduction stations with a real application»	2019	Risk analysis	[77]
«Simulation and health monitoring of a pressure regulating station»	2020	Health monitoring	[25]
«Machine learning methods for precise calculation of temperature drop during a throttling process»	2020	Health monitoring	[24]
«Natural Gas Hydrate Prediction and Prevention Methods of City Gate Stations»	2021	Data-driven prediction	[78]
«Risk analysis of gas leakage in gas pressure reduction station and its consequences: A case study for Zahedan»	2021	Risk analysis	[79]
«Natural Gas Consumption Forecasting Based on the Variability of External Meteorological Factors Using Machine Learning Algorithms»	2022	Load forecasting	[80]
«A deep reinforcement learning-based method for predictive management of demand response in natural gas pipeline networks»	2022	Demand-response	[27]

1.4 Aim

1.4.1 Motivation and research questions

Considering the importance of the continuous decarbonization of the NG transport chain and the safety issues concerning poor gas preheating, the limitations of the previous studies concern:

- the use of several lumped models for estimating pre-heating consumption that is simplified or not always compared with experimental data, or the choice of analyzing very complex and specific systems based on expanders or other technologies, is not always followed by a detailed analysis of the economic feasibility of the chosen system. If we were to divide a techno-economic analysis into two main parts, most articles studying this topic do not examine the two aspects comprehensively and focus on one of them at most. Furthermore, most of the plants present in Italy, but also in Europe, belong to the operational functioning classes (in terms, for example, of processed flow rate 20 Introduction and pressure drop) of medium-small size; consequently, it is very important to study also small-sized plants and the technologies that can be coupled to them and not only large and very large plants which, although characterized by high flow rates and therefore consumption, are much less widespread and therefore have a relative impact.

- the lack of reliable baseline consumption models of CGS operation that could be used to monitor

the operation of the thermal station and to assess the effects of possible modernization or efficiency-enhancing installations. While there are some studies, albeit few, on monitoring or of the state of health through data-driven methods or on the reliability of these systems, no one has raised the problem of actually going to evaluate the impact of the various technologies analyzed and proposed to enhance the efficiency of the preheating systems of the CGS or eventually replace them completely.

Considering the above mentioned gaps and problems, the research questions that this thesis will try to answer to are the following:

- This work proposes to develop a simplified, yet refined and general thermodynamic model to analyse the reduction of thermal energy consumption in a CGS equipped with RES-based air-source heat pumps, exploiting knowledge of the actual OCs of a CGS. Based on previous studies analysing CGS of comparable size, the authors focus on the inclusion of heat pumps for efficient gas pre-heating, exploiting dataset from several CGSs located in central Italy. The resulting research questions are:

  What kind of studies can be conducted with the simplified models in the literature and whether they can be improved?

  What are the optimal sizes of heat pumps from a techno-economic point of view to be coupled to these plants?

  Is it possible to achieve complete decarbonization of these systems while maintaining economic feasibility and acceptable resilience to energy price fluctuations?

- The second goal is therefore to develop a methodology to assess the best operating conditions of CGS from an energy-saving point of view, monitor energy performance and possibly propose counteractions or improvements. This work attempts to fill this knowledge gap and its aim is to develop a methodology of a CGS data set analysis and features selection to deploy several machine learning models, in particular regression models whose target will be the CGS preheating consumption, that will serve as a tool for monitoring tasks of the performance of the CGS preheating energy consumption. The resulting research questions are:

  What are the actual variables that influence the preheating requirements? What are the effects of seasonality on these plants? Is it possible to generalize the behavior of these systems?

  What types of relationships are there between consumption (the target) and input variables (flow rates, etc...)? How complex does the machine learning model have to be predictive and robust enough in these applications?

  Is it possible to develop an energy monitoring method for these applications, starting from energy baseline models?

1.4.2 Methodology

Figure 1.21 clarifies the methodology pursued to answer the research questions described in the previous section and thus realise the objective of the thesis work. Firstly, a set of super-simplified first-principles models will be developed to evaluate the preheating requests of 7 CGS whose annual datasets were provided by the DSO. Coupled with the CGS models, various RES-based tech models or tools were added for the evaluation of energy efficiency interventions within a CGS. Air source heat pumps are the key technologies considered for this type of analysis. These tools are applied to the techno-economic feasibility analysis of coupling conventional systems with renewable energy-based systems to reduce and possibly lower the carbon footprint of CGS. The study, therefore, turned to the realization of baseline

models of the energy consumption of these plants mainly for two reasons: the structural deficiency in realizing validated and precise physical models that can be generalized for this type of application and the possibility of accessing a lot of data provided by the DSO, which also made it possible to assess which are the real parameters that influence the consumption of these plants. The data-driven approach will be followed in the following steps involving the prototyping of models that replicate the behavior of the plants described above: analysis of the datasets provided by the DSO, realization of correlation studies between the variables and features engineering, training, and validation of regression models that act as a consumption 'baseline' with respect to the training period, and finally, use of these to carry out detailed energy monitoring analyses using CuSuM techniques. All models were developed in Python using some of the libraries available to the community, such as Scikit-Learn [81] for building the machine learning algorithms or CoolProp [82] for calculating the thermodynamic properties of fluids, in this case natural gas. Other parts of the tools were developed in-house in order to make up for the lack of models already in the literature, such as the thermodynamic preheating model.

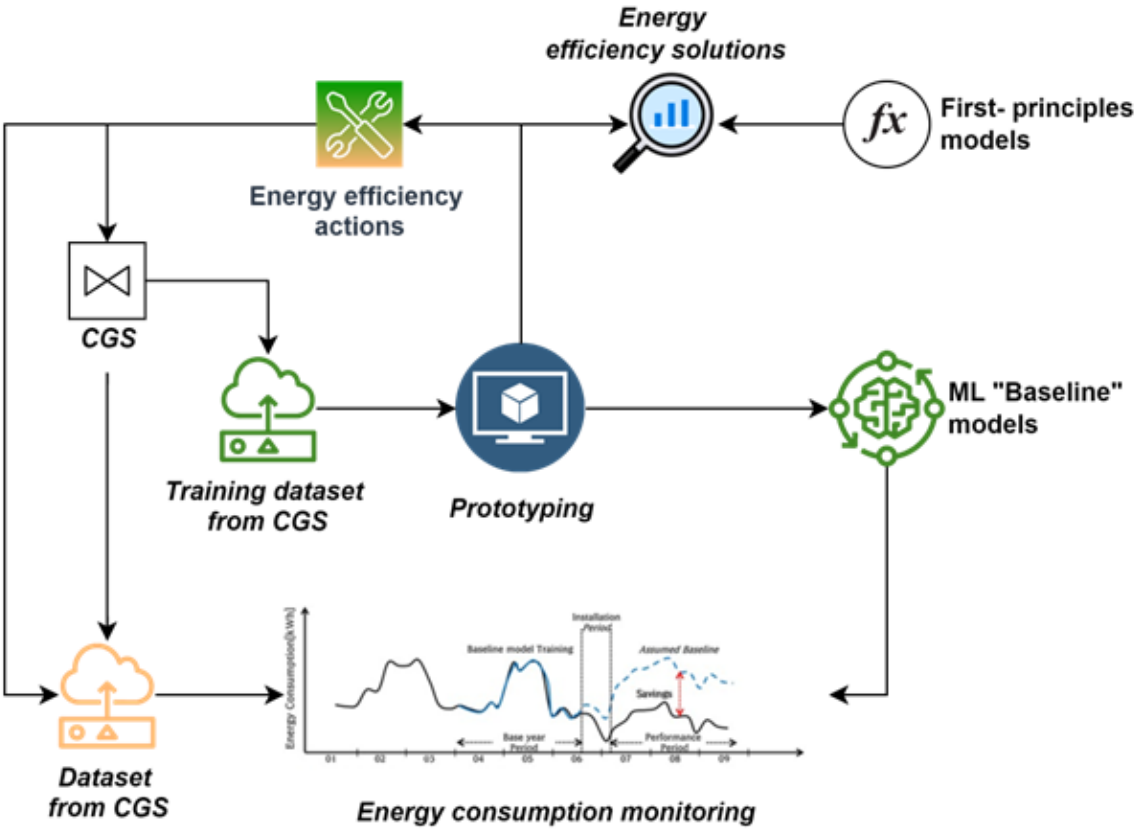


Figure 1.21: Thesis outline and methodology

Chapter 2

Materials and methods: decarbonisation of gas preheating

This section outlines the tools that will be used for the techno-economic feasibility analysis to decarbonize CGS plants equipped with renewable-based preheat heat generation systems. First, the proposed control logic with the various plant layouts is described, then the various sub-models, for preheating requirements, PV production, and HP output calculation are presented, and finally, the economic indices and references to energy prices and energy efficiency certificates are described.

2.1 CGS proposed hybrid layout

A new layout for the NG preheating system is proposed, based on the integration of a RES-based Heat Pump (HP), supported by a small Water Tank (WT), with the conventional gas-fired Boiler Unit (BU). The system is equipped with an Air Source Heat Pump (ASHP) fed by a Photovoltaic (PV) field, as it's shown in Figure 2.1. The system is also connected to the national grid and can either sell electricity only or be bi-directionally connected and buy it to power the electric heat pump. The various components of the CGS described in the figure will be explained in the following paragraphs. The assumptions underlying the operation of the HP are to utilise the outside air as a thermal reference well and to send water at a temperature of about 55 °C to the Preheater (PH), to preheat the gas before it enters the Throttling Valve (TV). This assumption is reasonable to maintain a safety coefficient at the heat exchanger to avoid flow crossings at any time of the year and at the same time not to penalise the efficiency of the heat pump too much. The flow rate of the preheating water is not calculated, since it is assumed that the system is able to modulate it with an inverter controlling the pumps speed, to manage the heat to be supplied to the gas optimally. In the following table, see 2.1, the assumptions are summarized with the respective symbols and values used in this chapter. There are three configurations that will be analysed in this chapter: the first, called off-grid, which only involves the installation of the HP with an auxiliary hot WT without any connection whatsoever to the grid, the second which only allows the PV electricity surplus to be sold, and the third and last which also allows purchases from the grid.

2.1.1 Hybrid system control logics

Every timestep Δt, assumed equal to 1 hour, the power balance between the pre-heating thermal demand of natural gas W_{PH}, for the given input conditions and output set points, and the heat output that can

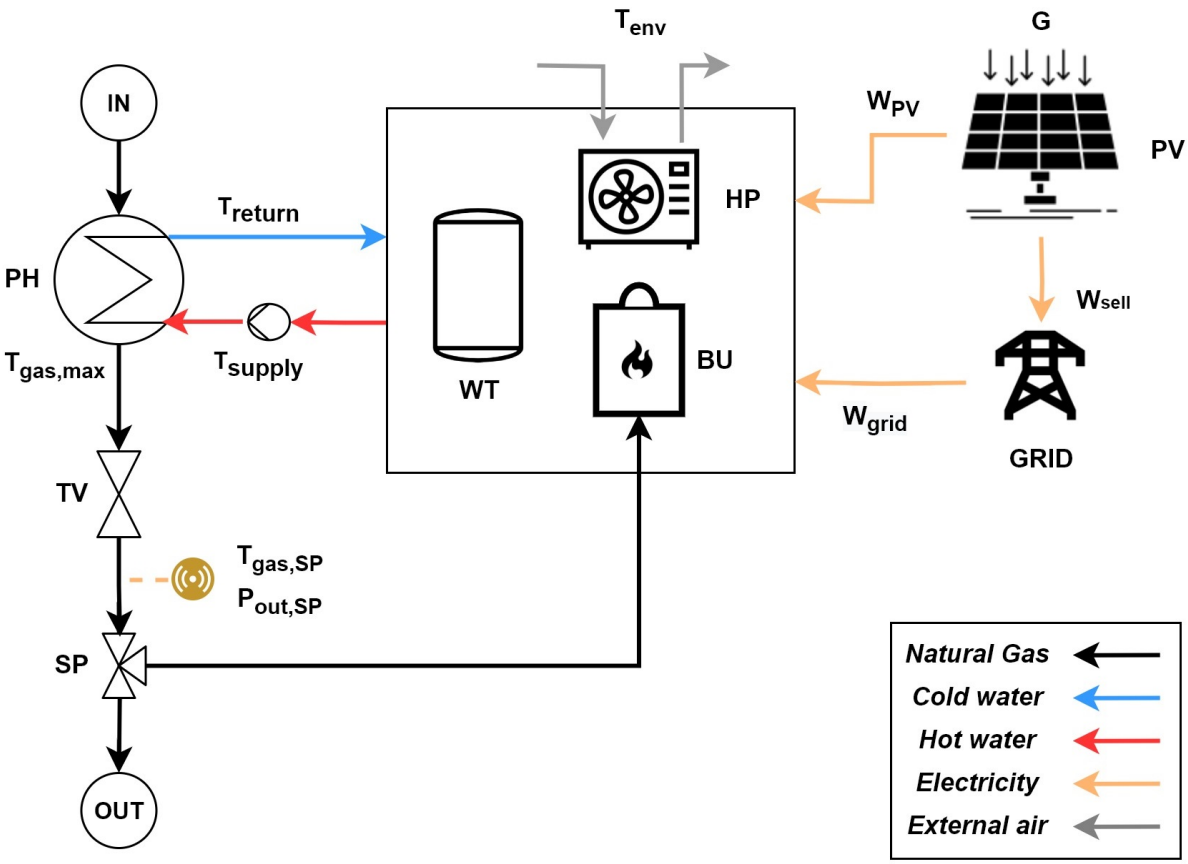


Figure 2.1: Natural gas preheating hybrid system layout

Table 2.1: Technical assumption for hybrid system operating conditions and models

Symbol	Name	Hypotesis
COP	Variable HP coefficient of performance	$COP = f(\Delta T_{lift})$
CF	Boiler capacity factor	$CF = f(W_{BU,max}, W_{BU})$
η_{BU}	Variable boiler efficiency	$\eta_{BU} = f(CF)$
η_{PH}	Constant preheater efficiency	$\eta_{PH} = 0.95$
T_{supply}	Supply water temperature constant	$T_{supply} = 55$ °C
$T_{gas,in}$	Gas inlet temperature dependency with air temperature	$T_{gas,in} = f(T_{env})$
$T_{i,SP}$	SET POINT temperatures according to thermal season	$T_{i,SP} = f(winter, summer)$
W_{ph}	Preheating variables in Standard conditions	$W_{ph} = f(\rho_{st,gas}, cp_{st,gas}, LHV_{st,gas})$
W_{PV}	PVGis tool for PV panel output calculation	PVGis tool [83]
μ_{JT}	Joule-Thomson calculation library	REFPROP [84]
$W_{tank,losses}$	Perfect mixing tank	$W_{tank,losses} = f(UA_{tank}, \Delta T)$

be supplied by the heat pump $W_{HP,th}$ is calculated. The heat pump always has priority; if there is a surplus of renewable thermal energy, i.e., electric power from the PV, all the preheating requirements are fulfilled with the heat pump and the equivalent surplus electricity is sold to the grid, if the control logic allows it. On the other hand, if the heat output supplying the HP is zero or insufficient, the water tank

and if necessary the auxiliary boiler comes into action and the necessary NG flow rate is obtained from the primary flow via the Splitting Valve (SP). The control balance of the logic is described by (2.1).

$$\Delta W(t) = W_{PH}(t) - W_{th,HP}(t) \tag{2.1}$$

The electric power exploitable by the ASHP every hour is obtained following a control logic that compares the power output of the solar panels with a "cut-off" threshold: when the solar field output is equal or higher to this threshold, the heat pump is switched on, but when the value of the PV output falls below the threshold, the heat pump is switched off. The threshold works that if the demand is less than 20 % of the heat pump's producibility, i.e. the size, the pump shuts down; this was decided discussing with the DSO, and because heat pumps have very constant part load efficiency curves, but the efficiency drops sharply for very high part load values (usually from 20 percent to 0 % of the maximum load). The heat output that the heat pump can provide every hour will be given by the product of the available electrical power and the actual coefficient of performance COP (t), see (2.31).

$$W_{th,HP}(t) = COP(t) \times W_{el,HP}(t) \tag{2.2}$$

Depending on whether the system can produce renewable electricity from PV alone or can purchase it from the grid, the $W_{el,HP}$ equation will change. In general form, therefore, equation (2.3) will be:

$$W_{el,HP}(t) = W_{PV}(t) + W_{grid}(t) \tag{2.3}$$

Table 2.2 summarises the possible control logic configurations that will be analysed in this chapter for this thesis work. The possible logics are as follows:

- *Logic 1: off-grid*: the heat pump can only receive electricity from the PV, see (2.4), or recharge the hot water tank, so the pre-heating requirement can only be met by a combination of $W_{HP,PV}$, W_{TANK} or W_{BU}.

$$W_{HP,PV}(t) = COP(t) \times W_{el,PV}(t) \tag{2.4}$$

- *Logic 2: on-grid (mono-directional)*: the heat pump can only receive electricity from the PV or recharge the hot water tank, so the pre-heating requirement can only be met by a combination of $W_{HP,PV}$, W_{TANK} or W_{BU}. In addition to that, the system can sell the PV surplus, i.e. when the $W_{PH} < W_{HP,PV}$, to the grid as W_{sell} to increase the profitability of the investment.

$$W_{sell}(t) = (W_{HP,PV}(t) - W_{PH}(t))/COP(t) \tag{2.5}$$

- *Logic 3: on-grid (bi-directional)*:the heat pump receive electricity both from the PV and from the grid and can also recharge the hot water tank, whose storage energy may be used when the heat pump output is not sufficient, so the pre-heating requirement can be met by a combination of $W_{HP,PV}$, $W_{HP,grid}$, W_{TANK} or W_{BU}. Likewise the second logic, the system can sell the PV surplus to the grid as W_{sell}.

$$W_{HP,grid}(t) = (W_{PH}(t) - W_{HP,PV}(t) - W_{TANK}(t))/COP(t) \tag{2.6}$$

If the $\Delta W(t)$ in the (2.1) is less than zero, i.e. we are in a thermal deficit condition even using the mains as a backup of electrical power, we are in a case where the pre-heating power is also greater than

Table 2.2: Powers involved in the balance of the various control logics

	W_{PV}	W_{tank}	W_{sell}	W_{grid}
Logic 1: off-grid	■	■	□	□
Logic 2: on-grid (mono-directional)	■	■	■	□
Logic 3: on-grid (bi-directional)	■	■	■	■

the maximum size of the HP plus the energy stored in the WT and therefore the boiler is switched on. For those scenarios, the power to be delivered by the boiler unit is given by (2.7:)

$$W_{BU}(t) = \Delta W(t) - W_{TANK} \tag{2.7}$$

The annual thermal energy demand request for the NG preheating process $E_{gas,y}$ (kWh/year) is obtained from (2.8), assuming the value of timestamp Δt =1h for a constant thermal power requirement every hour.

$$E_{gas,y} = \sum_{i=1}^{N_{hours}} W_{PH,i}\,(t) \times \Delta\mathrm{t} \tag{2.8}$$

The annual total volume of NG that need to be burnt in the BU is obtained with the following (2.9), considering the same hypothesis of the annual thermal energy demand calculation. In order to change from power to flow rate and then to volume of natural gas, it will be necessary to know the value of the Lower Heating Value (LHV) of NG, which will be introduced and described in the next sections.

$$V_{gas,y} = \sum_{i=1}^{N_{hours}} Q_{BU,i}\,(t) \times \Delta\mathrm{t} \tag{2.9}$$

The total annual thermal energy saved will be the thermal energy supplied by the pump instead of or together with the auxiliary boiler and it is given by (2.10).

$$E_{th,save,y} = \sum_{i=1}^{N_{hours}} W_{th,HP,i}\,(t) \times \Delta\mathrm{t} \tag{2.10}$$

2.1.2 Control logics evaluation parameters

Speaking of RES systems based on PV plants, the main parameters to be considered to assess the self-sufficiency level of the system are the SSR (Self Sufficiency Ratio) and SCR (Self Consumption Ratio), generally defined as ratios of amounts of electricity [85]. These are two indices that are used to evaluate the performance of control logic in renewable energy-based systems, particularly solar PV systems. In this study, since annual demand is thermal, the SCR of the RES system is adjusted to the considered case study and defined as the ratio between the self-consumed thermal energy and the total yearly energy demand 2.8; these two parameters are obtained from the following equations (2.11 and 2.12).

$$SSR = \frac{E_{th,sav,y}}{E_{gas,y}} \tag{2.11}$$

The SCR, on the other hand, can be expressed as the ratio of the self-consumed electric energy and the total yearly energy production.

$$SCR = \frac{E_{PV,HP,y}}{E_{PV,y}} = \frac{E_{PV,HP,y}}{E_{PV,HP,y} + E_{PV,grid,y}} \tag{2.12}$$

Where $E_{PV,HP,y}$ is the percentage of electricity produced by the PV throughout the year that is actually converted in heat with the HP and it is obtained directly from the knowing the actual COP value every timestep according to environmental conditions, see (2.4), $E_{PV,y}$ is the annual PV electric energy output and $E_{PV,grid,y}$ is the annual electricity sold to the grid.

2.2 Models

2.2.1 CGS preheating station

The thermal power W_{PH} required to preheat the standard gas flow rate $Q_{st,gas}$ before it enters the throttling valve is given by the equation below:

$$W_{PH} = \frac{\dot{Q}_{st,gas} \cdot \rho_{st,gas} \cdot c_{p,st} \cdot \Delta T_{gas}}{\eta_{ph}} \tag{2.13}$$

Where $\rho_{st,gas}$ is the NG density, $c_{p,st}$ is the specific heat capacity, both evaluated according to standard conditions (P = 1 atm, T = 15 °C) [5], ΔT_{gas} is the gas temperature increasing and η_{ph} is the preheating system efficiency, equal to 0.95 [20]. Henceforth, all volume or flow rate definitions will be expressed in standard cubic metres (Sm^3). The NG thermodynamic properties in general are assumed constant and equal for example to $\rho_{st,gas} = 0.763$ kg/Sm^3 and to $c_{p,st} = 2.160$ kJ/kgK, according to the annual values given by the Italian TSO. For this thesis work, thermodynamic properties are calculated using the CoolProp open source library database [82], which relies on the Nist RefProp libraries [84]. The following three quantities are then calculated knowing the standard conditions and concentrations of the chemical species (see Table 2.3) of the gas and passed to the software which returns the value of these:

$$LHV_{st,gas} = LHV_{gas}(P_{st}, T_{st}, [X_1, X_2, ..., X_n]) \tag{2.14}$$

$$c_{p,st} = c_p(P_{st}, T_{st}, [X_1, X_2, ..., X_n]) \tag{2.15}$$

$$\rho_{st,gas} = \rho_{gas}(P_{st}, T_{st}, [X_1, X_2, ..., X_n]) \tag{2.16}$$

Table 2.3: Comparison between different natural gas origins: mixture composition percentage [84]

Gas Name	CH_4	N_2	CO_2	C_2H_6	C_3H_8	Iso-Butane	C_4H_{10}	Iso-Pentane	C_5H_{12}	C_6H_{14}
CH4	100	0	0	0	0	0	0	0	0	0
Amarillo	90.6	3.12	0.46	4.53	0.83	0.103	0.156	0.032	0.044	0.039
Gulf Coast	96.5	0.26	0.60	1.82	0.46	0.098	0.101	0.0473	0.032	0.066
Ekofisk	85.9	1.01	1.50	8.50	2.30	0.349	0.351	0.051	0.048	0
Typical	95.1	0.09	2.56	1.84	0.24	0.040	0.016	0.014	0.011	0.08

In order to transit safely through the CGS, the gas must be preheated before the lamination step. The

required gas temperature increase is calculated using (2.17), where ΔT_{gas} is the sum of the difference between the gas inlet temperature and the gas outlet temperature $\Delta T_{out,in}$, given by (2.18), plus the gas temperature decreasing due the Joule-Thomson effect of (2.19).

$$\Delta T_{gas} = \Delta T_{out,in} + \Delta T_{JT} \tag{2.17}$$

$$\Delta T_{out,in} = T_{gas,out} - T_{gas,in} = T_{gas,SP} - T_{gas,in} \tag{2.18}$$

$$\Delta T_{JT} = \mu_{JT} \cdot \Delta P_{gas} \tag{2.19}$$

The temperature decreasing due the Joule-Thomson effect is given by 2.26, where μ_{JT} is the Joule-Thomson coefficient in °C/MPa and ΔP_{gas} is the pressure drop, which is calculated with the following equation.

$$\Delta P_{gas} = P_{gas,in} - P_{gas,out} = P_{gas,in} - P_{gas,SP} \tag{2.20}$$

Where $P_{gas,in}$ is the gas inlet pressure and $P_{gas,out}$ is the value of the gas outlet pressure, which is kept fixed in real and obtained giving the valve set point $P_{gas,SP}$.

The Joule-Thomson coefficient is generally assumed to be constant and equal to 0.4-0.5 °C/MPa in many studies [15, 16], however, for this thesis work, it is calculated for all different CGS conditions and for the type of gas chosen. To finally evaluate the NG flow rate to be taken from the main gas stream for feeding the Boiler Units Q_{BU}, the following equation is used:

$$Q_{BU} = \frac{W_{BU}}{LHV \cdot \eta_{BU}} \tag{2.21}$$

Where W_{BU} is given from the balance of control logic, see (2.1), η_{BU} is the Boiler Unit mean efficiency ($\backsim$ 0.85 or according to the Capacity Factor (CF) of the Boiler [20], as shown in Figure 2.2) and LHV is the Lower Heating Value of the Natural Gas, taken from the TSO database and depending on gas mixture origin.

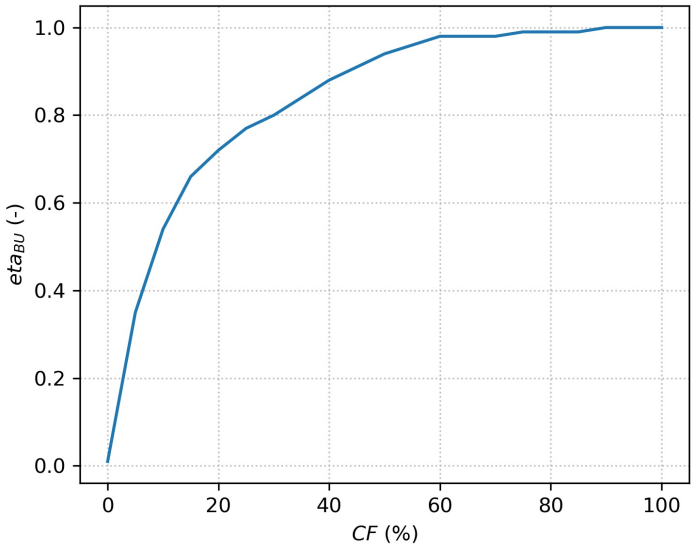


Figure 2.2: Boiler unit efficiency versus capacity factor according to [20]

It can be seen from this figure that the boiler tends to work at a very good efficiency and above 80% as long as the Capacity Factor CF is greater than 20%; for CF values below 20% the efficiency plummets and reaches zero for CF = 0%. CF is obtained from (2.22) as the ratio of the heat output required from the boiler to the maximum heat output that can be extracted from the BU.

$$CF = \frac{W_{BU}}{W_{BU,max}} \tag{2.22}$$

Gas inlet temperature models

The temperature of the gas arriving at the CGS is generally unknown, because of the lack of sensors placed at the station inlet, and that's the reason why is usually assumed constant in several scientific works and equal to the worst case possible, i.e. 5 or 0°C. In several other works, the authors exploit a more realistic model [13], which calculates the soil temperature surrounding a 1 meter depth buried pipe close to the CGS according to the variation of the air ambient temperature, assuming the NG temperature inside the pipes equal to the soil temperature, see (2.24) and (2.23).

$$T_{soil} = 0.0084\ T_{env}^2\ +\ 0.3182 \cdot T_{env}\ +\ 11.403 \tag{2.23}$$

$$T_{gas,in} = T_{soil} \tag{2.24}$$

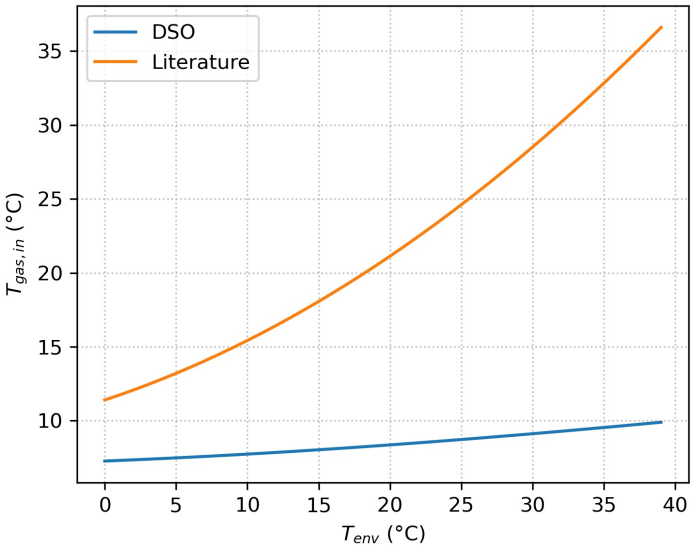


Figure 2.3: Natural gas inlet temperature vs air ambient temperature according to DSO (in blue) and to literature [58] (in orange)

This model proposed in the literature, however, when compared with the experimental data provided by our partner DSO, tends to greatly overestimate the incoming gas temperature for medium to high ambient temperatures. Consequently, in order to overcome this problem, it was decided to use a model obtained by interpolating the data provided by the DSO relating to a measurement of the inlet temperature at one plant, in particular CGS_7 among those in the cluster of cabins provided. The equation is the following:

$$T_{gas,in} = -0.000004 \cdot (T_{env}^3) + 0.0009 \cdot (T_{env}^2) + 0.0382 \cdot (T_{env}) + 7.2727 \tag{2.25}$$

The data collected by the DSO to generate the second correlation curve are not shown for confidentiality

reasons. The limits of these two models are such that when the parabolic curve reaches the minimum, that is the point at which the curve would reverse giving rise to nonphysical behavior (inlet gas temperature increasing as the ambient temperature decreases), and therefore it is necessary that the value from the above condition be constant and was obtained equal to T_{env} = -18°C for the first model and T_{env} = -7°C for the DSO model.

Joule-Thomson coefficient

The Joule-Thomson effect is a phenomenon whereby the temperature of a real gas decreases following expansion conducted at constant enthalpy. In literature, this parameter is often assumed constant and equal to 4-5 °C/MPa during the gas throttling process inside a CGS or its calculation is avoided by imposing an isenthalpic transformation between the starting point and the end point. For this work purpose the authors calculate the coefficient according to the following formula:

$$\mu_{JT} = \left(\frac{\partial T}{\partial P}\right)_H \tag{2.26}$$

The Joule Thomson coefficient μ_{JT} is calculated for several NG mixtures, stored in the Cool Prop database [82], which are enlisted and described in Table 1 and for customized mixtures that represent the Italian Scenario. The gas outlet condition (P_{out}, T_{out}) is kept fixed an equal to the ideal set point values for pressure and several set point temperatures (3 bar, 10 °C) to compute the isenthalpic process necessary for the μ_{JT} evaluation. Figure 2.4 shows the linear dependence between the Joule-Thomson coefficient, which approximately varies between 4.1 °C/MPa and 5.2 °C/MPa, and the inlet pressure for all the considered NG mixtures. The pure methane (CH_4) turns out to be the fluid with the lowest value of μ_{JT}, while the Ekofisk (North European) NG is the mixture with the highest temperature drop during the throttling phase at constant enthalpy. The Typical NG μ_{JT} curve is chosen to be the one that will be used in the following chapters, to consider a NG mixture composition more general as possible, since in Italy there is a very high variability of gas composition due to the multitude of import origins. Focusing on the typical NG μ-curve, Figure 2.5 highlights the effect of a different output temperature set point on the μ_{JT} coefficient and thus on the final preheating energy demand. Increasing the output temperature set point causes the coefficient to decrease with the same input conditions. As the gas outlet set point increases, the temperature drops to be made by the gas at the same inlet pressure at the CGS are reduced, and thus the share calculated with (2.18). On the other hand, a higher set point leads to an increase in the preheating factor calculated with (2.19). Thus, the gas outlet set point will affect the two components of the overall temperature drop equation in opposite ways.

.

Real operating conditions

In this section, a real dataset covering one year of operation of two plants, called CGS_A and CGS_B, located in central Tuscany, will be exploited as a comparison to the model results. The gas outlet temperature $T_{gas,out}$ is the key parameter to be monitored and it is strictly dependent on the value of the outlet temperature Set Point which is set by the DSO inside all the CGS. According to the Italian framework, this value must be at least equal or above 5 °C , but for safety reason is generally set at higher values and it can be modulated for two types of working seasons: winter, and summer. In this work two different values for the outlet gas temperature are considered, replicating the actual output temperature setting inside a CGS in Italy, as it can be seen in the following equation:

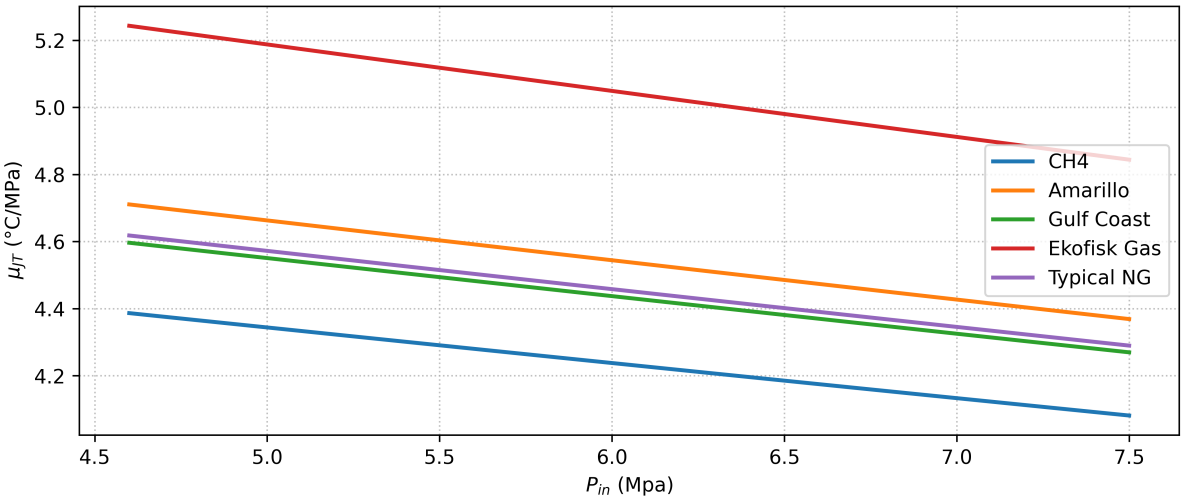


Figure 2.4: Joule-Thomson coefficient μ_{JT} for several gas mixtures vs the inlet pressure P_{in}

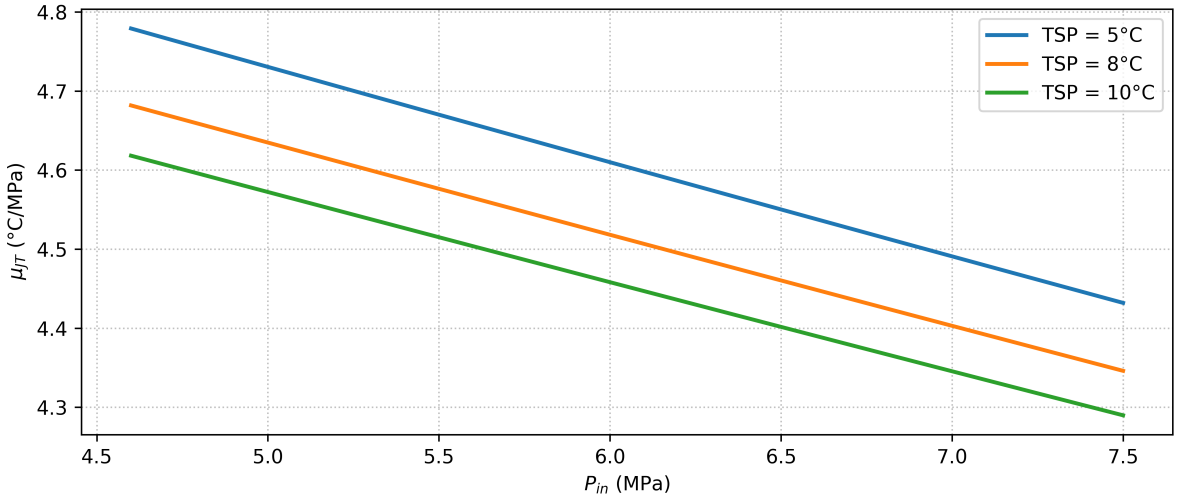


Figure 2.5: Joule-Thomson coefficient μ_{JT} for the natural gas typical composition for several gas outlet temperature set points vs the inlet pressure P_{in}

$$T_{gas,out} = T_{gas,SP} = \begin{cases} 8\ C & \text{if } \textit{winter thermal season} \\ 10\ C & \text{if } \textit{summer thermal season} \end{cases} \tag{2.27}$$

Another important hypothesis concerns the control logic of the preheating system: to generalise the work as much as possible and after talking with the DSO, it was decided to follow a logic based on modulation of the water flow rate according to the heat to be supplied to the gas. Once the water flow temperature to the PH is fixed and the gas conditions are known, the flow rate is derived accordingly to perfectly match the thermal demand at PH.

The model used to calculate preheating gas consumption was refined by adding improving assumptions if compared to similar 0-D models used in literature, see for example [15, 16, 20]. The hypotheses that have been added gradually are as follows:

- Hp1: Variable gas flow rate crossing the CGS instead of a single constant value.
- Hp2: Real outlet temperature set point according to DSO.
- Hp3: Variable gas inlet pressure

- Hp4: Variable gas inlet temperature depending on ambient air temperature with model of (2.25)

Table 2.4: Description of all and hypotheses that were added to the model successively

Model ID	Q_{gas}	$T_{gas,out}$	P_{in}	$T_{gas,in}$
BL (Base Load)	Design	Max (10 °C)	Max (7.5 MPa)	Min (0 °C)
BL + Hp1	Real	Max (10 °C)	Max (7.5 MPa)	Min (0 °C)
BL + Hp1 , Hp2	Real	Real SP (10/8 °C)	Max (7.5 MPa)	Min (0 °C)
BL + Hp1 , Hp2 , Hp3	Real	Real SP (10/8 °C)	Real P_{in}	Min (0 °C)
Final Model	Real	Real SP (10/8 °C)	Real P_{in}	$T_{gas,in} = f(T_{env})$

Table 2.4 shows the starting point of the model, called Base Load, and the end point at which it could be reached by adding improving assumptions called HP_i. The first assumption that was added was to consider a variable gas flow rate instead of a constant design flow rate, the second that of adding realistic outlet gas set points, the third to use values of real inlet and outlet gas pressures and not jumps in design pressure, and finally to add the inlet gas temperature model instead of evaluating only a lower boundary temperature, such as for example 5°C or 0°C, as is done in several scientific papers in the literature. Figure 2.6 shows the effect of including each hypothesis on the final result. Adding a real flow rate changes the estimate from about more than 600% from the final value obtained with the final model to about 160%. You then get less pronounced but still important improvements by adding seasonally varying set points, more accurate inlet pressure and inlet temperature values.

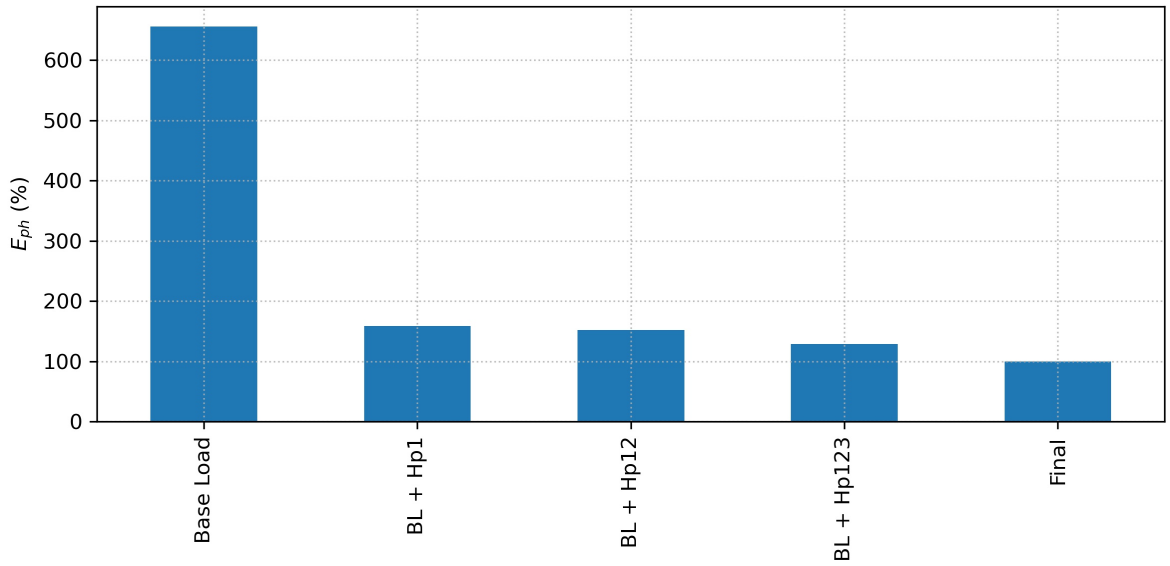


Figure 2.6: Preheating requirements with different hypotesis showed in Table 2.4.

Preheating requirements model errors

Models results have been compared with real data related to the one-year operation of CGS_A and CGS_B. Figures 2.7 and 2.8 show the percentage error of the model prediction, throughout the year. The percentage error was calculated as the ratio between the difference of the model output with the real value and the real value itself for each timesample, through the following equation:

$$e = (W_{PH,model} - W_{PH,real}) * 100/W_{PH,real} \tag{2.28}$$

Error-values as high as 400% of the actual value are observed, indicating that the model prediction is 4 times higher than the measured consumption. The percentage of errors is very high, especially for the summer and intermediate seasons. If, on the other hand, only the winter season is considered, as shown in Figures 2.9 and 2.10, the percentage error is greatly reduced, ranging from +40% to -60%. For both plants, the model underestimated the thermal energy demand by up to 60 percent for CGS_A and about 42 % for CGS_B. This is because in summer there is no demand for preheating, and the model tends to overestimate the prediction with very high percentage errors but very low absolute values (less than 0.6 Smc/h versus pilot flame consumption estimated to be around 0.25 Smc/h per boiler). Considering winter, where preheating consumption is relevant, the percentage error is reduced to orders of magnitude around 50 percent. In this case, it is not possible to estimate individual plant losses, losses in the circuits, and any inefficiencies in the plant's operation.

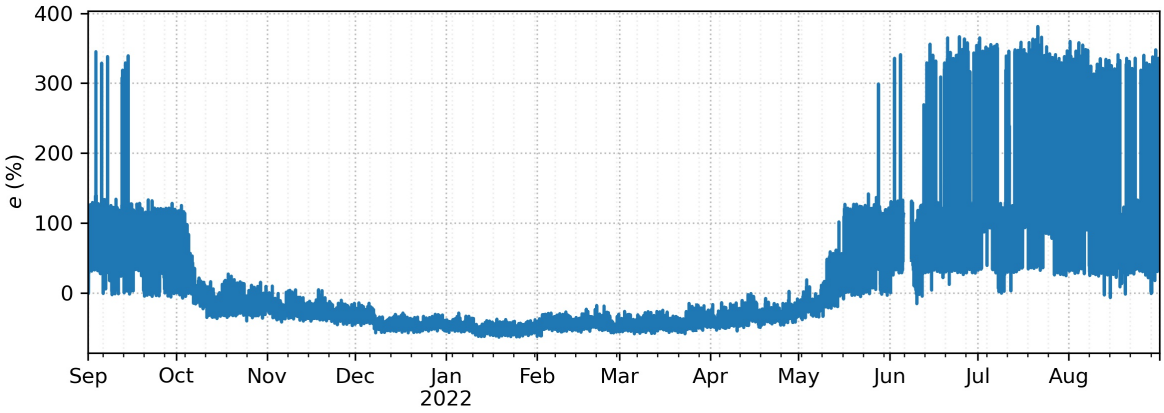


Figure 2.7: Percentage error of the preheating model output respect to the real data for the CGS_A

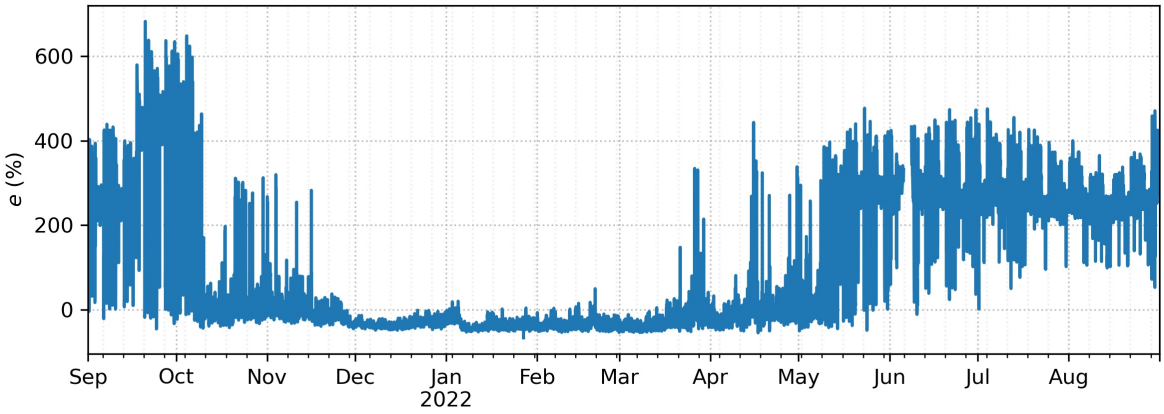


Figure 2.8: Percentage error of the preheating model output respect to the real data for the CGS_B

Figures 2.11 present the results of comparing hourly trends of heat load power calculated with the model during the year and during 4 typical days related to operation of CGS_A, respectively. The demand curves of the model and real data are scaled dimensionally to make the treatment generic for any CGS: the model with real inputs can faithfully replicate the consumption trends for all seasons of the year in qualitative terms, as can be seen in the above figure. Figure 2.11 shows that the model is very accurate

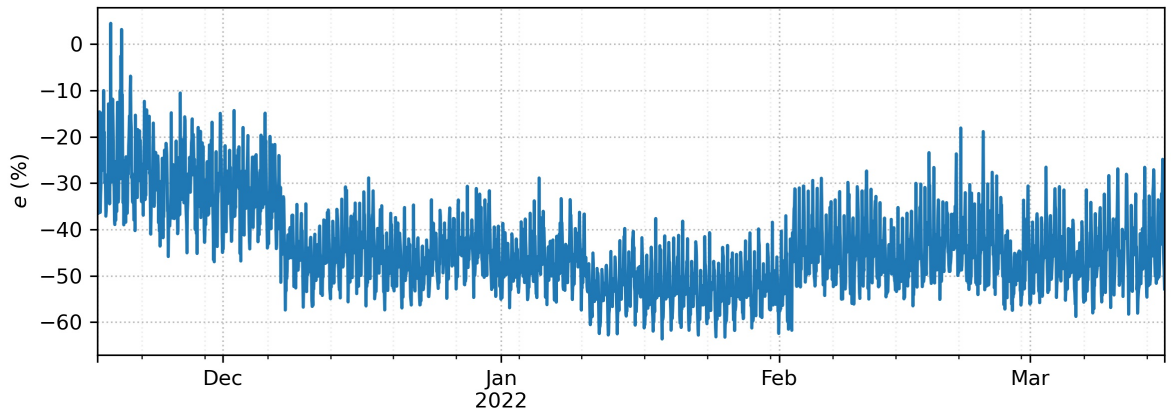


Figure 2.9: Percentage error of the preheating model output respect to the real data for the CGS_A, considering the winter season

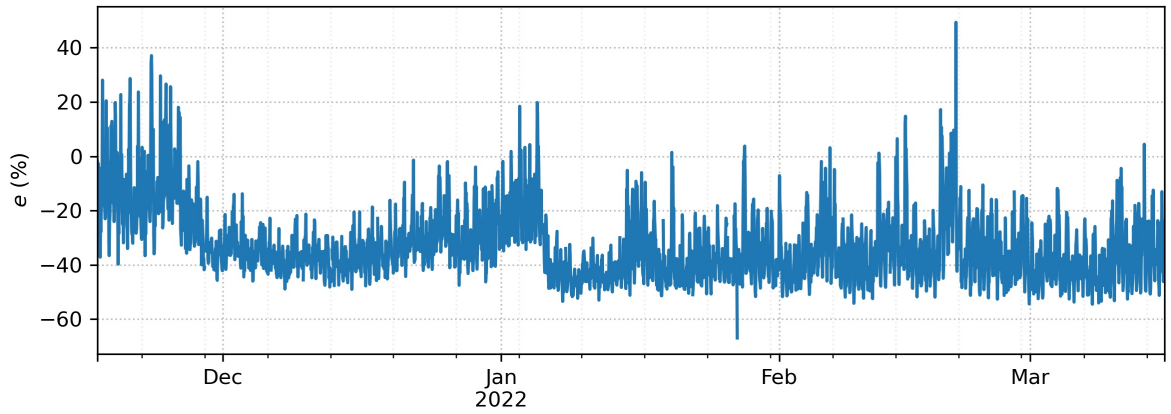


Figure 2.10: Percentage error of the preheating model output respect to the real data for the CGS_B, considering the winter season

when the system is operating at maximum load (daytime peak hours and winter seasons). On the other hand, the model tends to underestimate or overestimate at other times of the day, particularly at night and in summer, when the gas flow rate is very low and the system presumably retains some amount of heat loss. It is clear from the figures that the operation of the cabin is strongly influenced by the OCs in which it operates: when it is at maximum load, the relationship of (2.13) between consumption W_{PH} and gas flow rate Q_{gas}, with all the assumptions added during model development, is optimal.

The results of the comparisons of the model just described with actual values are reported directly in this section and not in the Results section, since validation of the model is beyond the scope of this thesis. It was not possible to validate the previously described model for the 2 plants considered. The plants in question are the only ones that have access to a data-set of preheating consumption among all the plants we will consider for this work which will be described later in Chapter 4. It is very difficult to validate an instrument based on super-simplified equations with actual data recorded by the instruments of the plants that the DSO provided us with the data. Validation of these instruments would require an effort that for techno-economic study and scenario analysis would not be justified by the final result, for two main reasons: the variety of plants in terms of power sizes and plant layouts is an obstacle to using a simple model in a massive and generalized way and the difficulty of including thermal circuit losses for each plant in a single model based on thermodynamic equations.In any case, our model managed to perform significantly better (up to 600% more precise if compared to real data), analogized to a type of treatment

often used in the literature, i.e. assuming the various operating parameters of the system constant and with very conservative coefficients. To achieve sufficient accuracy to faithfully represent the plant and also in light of these results, we defer to the next conceptual step followed by this thesis: the development of machine learning models through data-driven approaches, which would allow for accurate digital twins without requiring excessive and unwarranted validation efforts.

2.2.2 Techs

PV model

To evaluate the solar PV panel production nearby the several CGS, the European Commission tool PVGIS is used. PVGIS stands for Photovoltaic Geographical Information System and provides information about solar radiation and photovoltaic (PV) system performance for any location in Europe and Africa, as well as a large part of Asia and America [83]. The panel power for each size is obtained by multiplying the dimensionless curve obtained based on the location of the CGS with the specific size considered, as described in (2.29).

$$W_{PV}(t) = W_{PV,1kW} \times PV_{size} \tag{2.29}$$

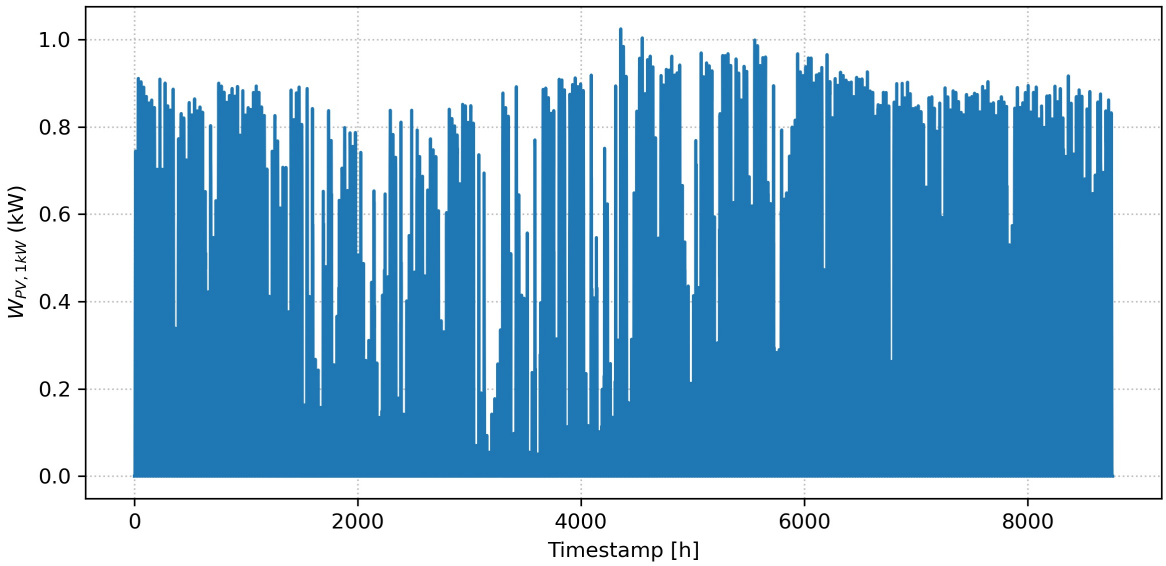


Figure 2.12: PV production output example for a 1 KWp solar panel using PVGis tool.

Heat pump model

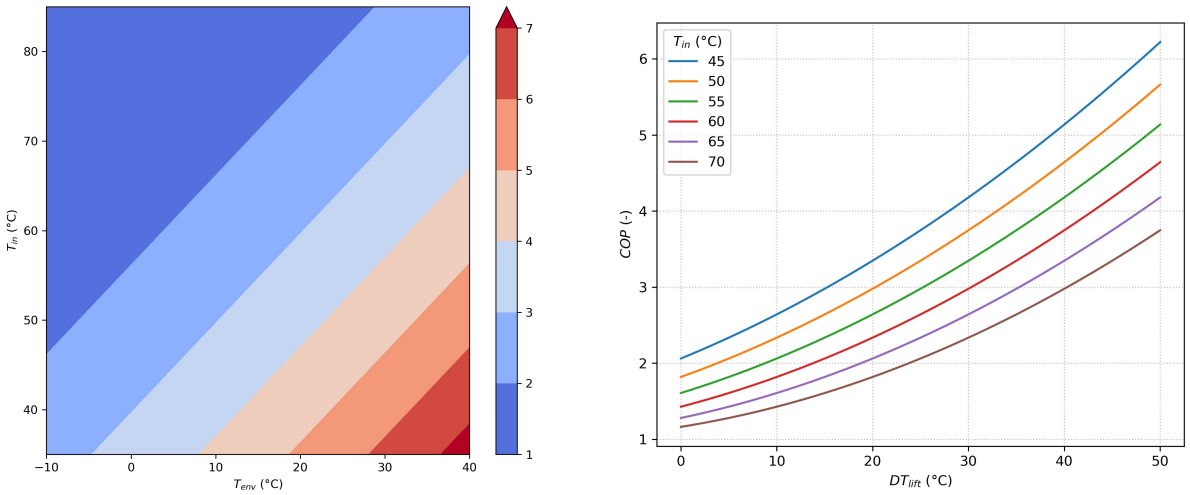
The HP is a very efficient technology for heating and cooling purposes, since its efficiency, the COP (Coefficient of Performance), usually varies from 2 to 5 and is particularly high when used to heat a utility or process. The value of the COP is defined according to 2.30 as the ratio of the thermal output with the electricity input but for this work will be computed using 2.31, considering the efficiency dependence on the temperature difference between the water supply temperature T_{in} and the ambient air temperature ΔT_{lift} [86], describe in (2.32).

$$COP = W_{th,HP}/W_{el,HP} \tag{2.30}$$

Of the various types of heat pumps, it was decided to focus on just one, and in particular on the most

commercially popular, the ASHP (Air Source Heat Pump); the advantages are that it is easy to install and works as a heat sink with the surrounding air. The trend of the COP of the HP chosen for this work is plotted in 2.13, which follows the model of (2.31).

$$COP_{ref}\,(t) = 6.81 \;-\; 0.121 \cdot \Delta T_{lift}(t) \;+\; 0.00063 \cdot \Delta T_{lift}(t)^{\,2} \tag{2.31}$$



(a) (b)

Figure 2.13: (a) COP for ASHP according to Kozarcanin et. al. for different values of T_{in} and environmental temperature T_{env} vs (b) COP map for different T_{in} vs ΔT_{lift} [86].

The equation shows a quasi-linear trend as the ΔT_{lift} varies and consequently the external temperature T_{env}, for the same delivery temperature T_{in}. Very high values of COP could be reached if the outside air temperature is very high and the supply temperature is reduced to values close to the environmental temperature ones. In that case you may reach values up to 7 of the efficiency of the HP. However, the system will have mainly to work more in winter season when the preheating request is higher and the ambient temperature lower, and the delivery temperature will have to exceed a certain safety value, generally higher than the maximum temperature reachable by the gas exiting the heat exchanger, to avoid heat flow inversions in the preheater. Because of this combination of facts, the HP will generally work in the zone that varies from 2 to 4 for the COP. This can also be seen clearly from the second figure, the efficiency curve of the machine will decrease as the delivery temperature to the gas preheater increases.

$$\Delta T_{lift}(t) = T_{HP,out} - T_{env}(t) \tag{2.32}$$

The heat output that the heat pump can provide every hour will be given by the product of the available electrical power and the coefficient of performance under those operating conditions, i.e. as the outside air temperature varies.

$$W_{th,HP}(t) = COP(t) \times W_{el,HP}(t) \tag{2.33}$$

Storage water tank

The sizing of a small hot water tank is planned for the storage of the excess thermal energy produced by the heat pump. These systems do not allow for seasonal energy storage, to which a long-term storage system would be preferable, as is gaining ground lately through bidirectional hydrogen green fuel cell -

-electrolyzer systems, but they allow for the storage of a small amount of energy and discharge it to avoid using the boiler and reduce waste when there is a renewable surplus.

The assumptions upon which the tank sizing is based are:

- Tank temperature working range between $T_{tank,SP} = 60°C$ and $T_{tank,low} = 40°C$
- Maximum energy storage capacity: 50% of W_{PH} for 1 hour
- Water properties equal to $\rho_{H_2O} = 997$ kg/m3 and $cp_{H_2O} = 4.187$ kJ/kgK

In fact, the tank functions as an inertial buffer between the heat pump and the gas preheating circuit, only that compared to a domestic solution, the tank will always be undersized if we only consider the heat pump. The heat transfer coefficient U $[W/m^2K]$ is taken from [87] and it's used to compute the overall heat transfer coefficient UA [W/K] of (2.34). The area of the tank is calculated starting from the calculation of the V_{tank}, as it can be seen in (2.35).

$$UA_{tank} = U_{tank} \times A_{tank} \tag{2.34}$$

$$V_{tank} = \frac{E_{tank}}{cp_{H_2O} \times \rho_{H_2O} \times (T_{tank,SP} - T_{tank,low})} \tag{2.35}$$

Tank losses each time-step due to heat loss to the atmosphere are given by (2.36).

$$W_{losses} = (UA_{tank} \times (T_{tank,in} - T_{env})) \tag{2.36}$$

Once the losses are calculated, they are used to derive the tank temperature at the end of the timestamp. This temperature will then be the temperature T_{tank} (t+1) at the start of the next timestamp. The temperature is given henceforth by (2.37).

$$T_{tank}(t+1) = T_{tank}(t) + \frac{W_{tank}(t) - W_{losses}}{(m_{tank}/\Delta t) \times cp_{H_2O}} \tag{2.37}$$

Where Δt is the calculation timestamp equal to 1 hour, T_{tank} is the temperature of the previous timestamp, m_{tank} is the mass of the water contained in the tank [kg] and cp_{H_2O} is the specific heat of the water.

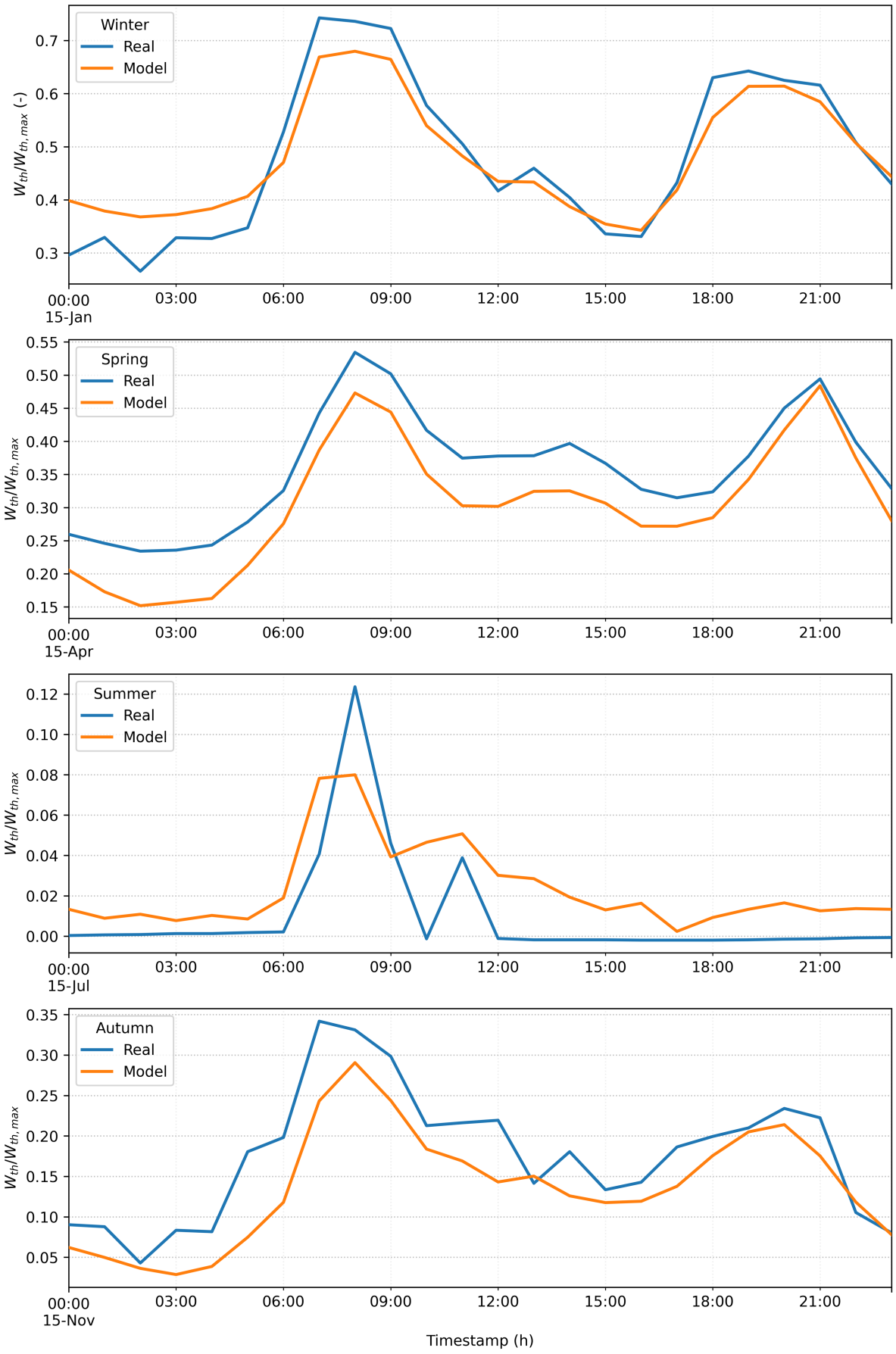


Figure 2.11: Daily comparison between the hourly required dimensionless thermal power: final model vs real data of CGS_1 for 4 different seasons

2.3 Economics

This section presents the economic indices used to calculate the techno-economic feasibility of the proposed plant solutions and the results of which will be shown in chapter 5. The Net Present Value (NPV) formula of the investment is applied for each individual configuration analysed and the results will be shown in the next section. The following assumptions were included in the formula (2.38):

- Interest rate i = 6 %.
- Number of years n = 20
- Overall cash flow rate given by the total amount of energy efficiency certificates EECtot for the number of years granted by the system operator [88] nEEC.

NPV formula:

$$NPV = \sum_{j=1}^{n} \frac{CF_y}{(1+i)^j} - I_0 + \begin{cases} EECtot & \text{if } j < nEEC \\ 0 & \text{if } j > nEEC \end{cases} \tag{2.38}$$

The yearly cash flow is given by (2.39): the saving of natural gas is a gain, together with the sale of renewable surplus to the electricity grid; on the other hand, expenditure items include the cost of purchasing from the electricity grid when allowed by the control logic and the cost of plant maintenance for all three logics.

$$CF_y = C_{NG} \times V_{save,y} + C_{el} \times E_{sell} - C_{O\&M} - C_{el,1} \times E_{grid} \tag{2.39}$$

Where $C_{el,1}$ is the purchased electricity price, C_{el} is the electricity selling price, $V_{save,y}$ is the amount of NG volume saved per year and E_{grid} and E_{sell} are the amount of energy purchased and sold to the grid, respectively.

Techs capital costs

The initial investment costs for technology acquisition are given by the following equations: (2.40), (2.41), (2.42) for the PV panels, the heat pump and the tank, respectively. The overall capex is then computed with (2.43) as the sum of the previous three economic components.

$$I_{PV} = C_{PV} \times W_{PV,peak} \tag{2.40}$$

$$I_{HP} = C_{HP} \cdot W_{HP,peak} \tag{2.41}$$

$$I_{tank} = C_{tank} \cdot E_{tank} \tag{2.42}$$

$$I_0 = I_{PV} + I_{HP} + I_{tank} \tag{2.43}$$

All prices for the photovoltaic panel investment cost were taken from the IRENA report of 2020 [89], while the capital cost of the heat pump from a more recent study on the decarbonisation of the NG sector in Italy [90]. The assumption is that the system is equipped with an inverter and that its price is included inside the capital cost and its useful life is 15 years, after which it must be replaced and its cost is around $c_{inv} = 5\%$ of the capital cost of installing the PV array, see (2.44).

$$I_{inv} = I_{PV} \times c_{inv} \tag{2.44}$$

Table 2.5: Techs capital costs.

	Unit	Cost (overall)	Reference
PV panel installation cost C_{PV}	€/kW_p	1200	[89]
Inverter replacement installation cost	%	5	[1]
ASHP installation cost C_{HP}	€/kW_{th}	$1855.2 \times HP_{size}^{-0.262}$	[90]
TANK installation cost C_{tank}	€/kWh_{th}	120	[91]

Techs operative costs

The investments operating costs are calculated with (2.45) and the values chosen are summarized in Table 2.6. The operative costs of the tank are neglected.

$$C_{O\&M} = C_{PV,O} \times I_{PV} + C_{HP,O} \times I_{HP} + C_{tank,O} \times I_{tank} \tag{2.45}$$

Table 2.6: System operative & maintenance costs.

	Unit	Specific cost	Reference
PV panel O&M $C_{PV,O}$	%	2	[1]
ASHP O&M $C_{HP,O}$	%	2.07	[90]
TANK O&M $C_{tank,O}$	%	0	[91]

2.3.1 Energy efficiency certificates and energy prices

The first step in the economic evaluation will be to calculate the number of Energy Efficiency Certificates (EECs) one can access based on the volume of natural gas saved in the year, expressed in tonnes of oil equivalent (TOE). A certificate is awarded for each TOE of natural gas, using the conversion between Sm^3 of NG and TOE, and approximating this value by default if the TOE unit is less than half, or by excess if equal or greater. For this work, a conversion factor of 0.836 TOE per 1000 Sm^3 of natural gas saved was chosen [5, 88]. Following the Inter-ministerial Decree of 11 January 2018 [88], the tariff contribution recognized for the White Certificate is €260 (the effective reimbursement of which is equal to €250) and is paid from 3 to 10 years depending on the type of intervention carried out. For this reason, the type of energy efficiency intervention that allows access to tax relief for 7 consecutive years at a price that was decided to set equal to 250 €, or the net gain of the DSO, was considered.

- Unitary revenue of energy efficiency certificate $EEC_i = 250€$.
- Number of years for the EECs equal to nEEC = 20

Electricity and gas prices have been recently fluctuating disproportionately compared to the pre-crisis and pre-war in Ukraine, and the fluctuations have been very wide and often unpredictable. As a price

baseline for calculating the NPV of the profit of each individual pair of technologies (HP + PV) for each individual plant, an average price was chosen for both gas and electricity, relating to the period used for the analysis.

Table 2.7: Energy prices

	Unit	Cost (operating year average)	Reference
Natural gas price	€/*Smc*	1.2	[88]
Electricity price (purchase)	€/*kWh*	0.35	[88]
Electricity price (sell)	€/*kWh*	0.05	[88]

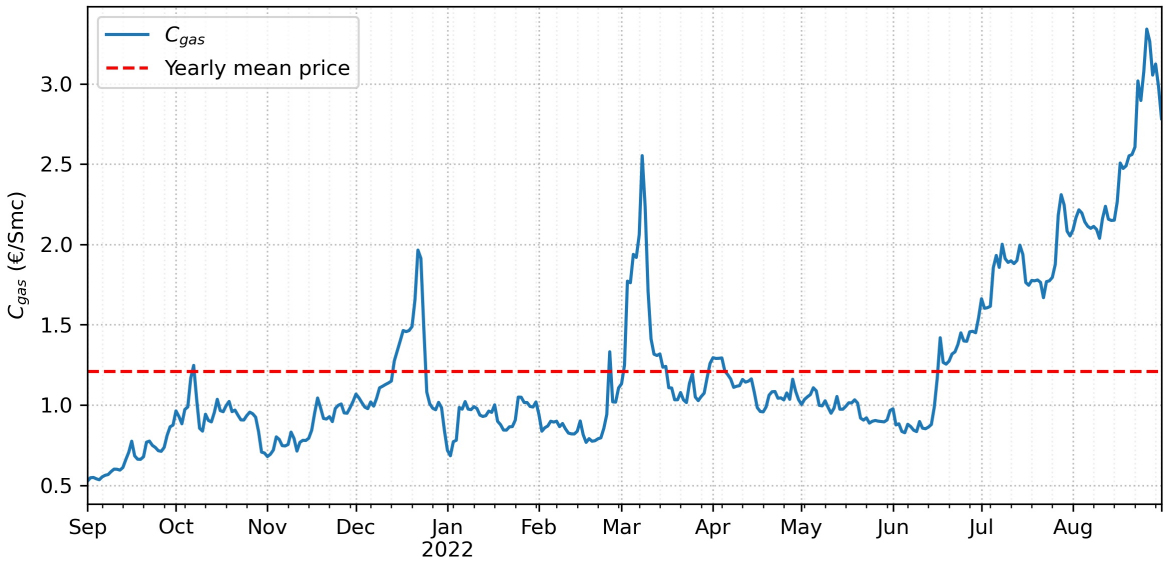


Figure 2.14: NG prices according to [92] for the considered period

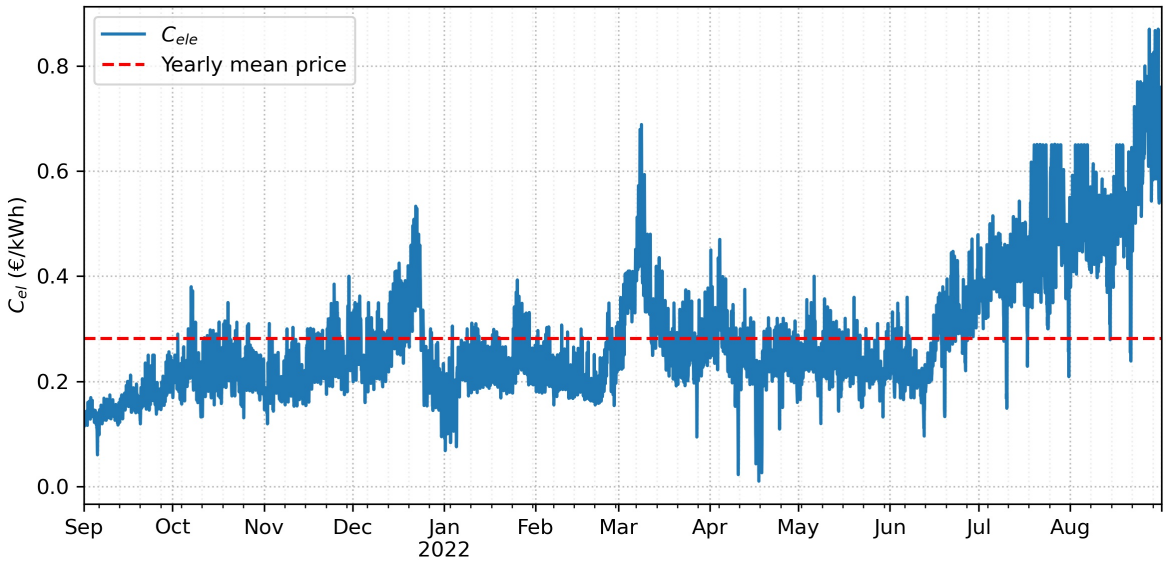


Figure 2.15: Electricity prices according to [92] for the considered period

The figure shows how the price of gas skyrocketed starting in September 2021 and reached peaks during

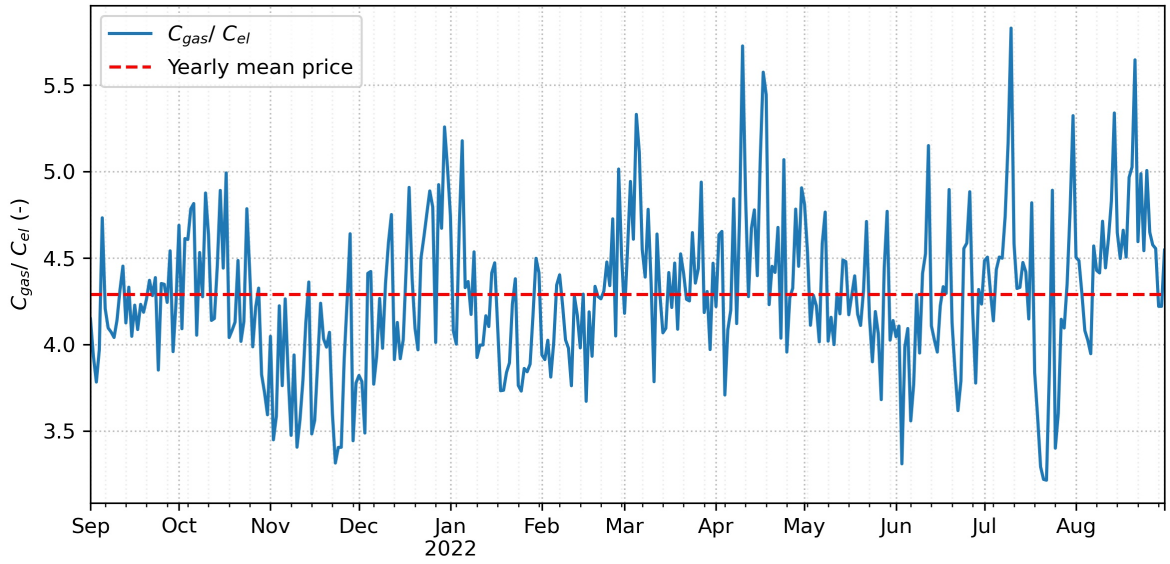


Figure 2.16: Prices ratio obtained from the ratio of the NG price and the electricity price for the considered period

the period when the war in Ukraine began, i.e., from February 2022, and after a period of readjustment returned to values even above €2/Smc. The average price for the period under consideration is about 1.2 €/Smc. The price of electricity has proportionally followed that of gas, in fact as can be seen from the figure 2.16, the ratio of the price of gas to that of electricity has fluctuated around an average value of about 4.2. This is certainly interesting for our analysis: it is not only the absolute value that influences the cost of technology and the gain that can be obtained from a given investment, but also the ratio of gas to electricity that is crucial. Since the objective of the work is to reduce gas consumption, therefore work on the price of the gas energy component, through 100% renewable and therefore free energy, or with energy developed from the grid, which has a cost comparable to that of gas, it will be important to keep focus on all these aspects at the same time. To answer this question, a sensitivity analysis was carried out which will be presented in the section of the results of the techno-economic analysis which will exploit the tools just presented.

Chapter 3

Materials and methods: machine learning-based energy monitoring

This chapter presents the method developed for data-driven modeling of preheat consumption in CGSs and describes all the tools needed to implement these models. Machine learning with its paradigms and approaches is briefly introduced, and then all the tools necessary for analyzing a dataset, choosing features, training and validating a model, and verifying its performance are described. Finally, the concepts of energy monitoring, baseline energy models, and the CuSUM method are introduced.

3.1 Machine learning: a brief introduction

Machine Learning (ML) is a subset of artificial intelligence (AI) that deals with creating systems that learn or improve performance based on the data they use. Artificial intelligence is a generic term and refers to systems or machines that mimic human intelligence. The terms machine learning and AI are often used together and interchangeably, but they do not have the same meaning. Deep Learning (DL) is that research field of machine learning and artificial intelligence that is based on different levels of representation, corresponding to hierarchies of characteristics of factors or concepts, where high-level concepts are defined based on low-level ones. In other words, deep learning means a set of techniques based on artificial neural networks organized in different layers, where each layer calculates the values for the next one so that the information is processed in an increasingly complete manner. Algorithms are the engines that power machine learning. The two main types of machine learning algorithms currently used are: supervised machine learning and unsupervised learning. The difference between these two types is defined by the way each algorithm learns data to make predictions. There is also a third category called reinforcement learning.

- Supervised machine learning: it is defined by its use of a labeled data-set to train algorithms that to classify data or predict outcomes accurately. As input data is fed into the model, it adjusts its weights until the model has been fitted appropriately, which occurs as part of the cross-validation process. Linear and logistic regression algorithms, multi-class classification and support vector machines are some examples of supervised machine learning.

- Unsupervised machine learning: it uses a more independent approach, in which a computer learns to identify complex processes and patterns without the careful and constant guidance of a person. Unsupervised machine learning involves training based on unlabelled data for which no specific

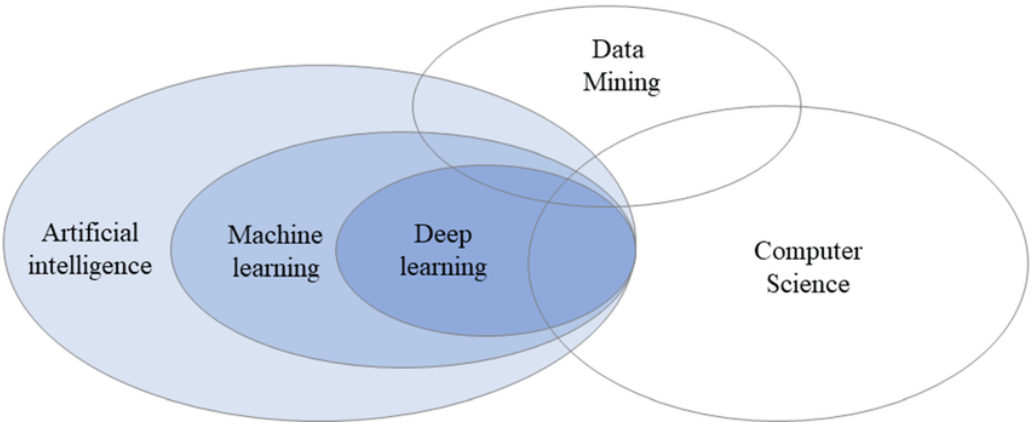


Figure 3.1: Venn diafram represening the relationship between AI-ML and DL [93]

output has been defined. The k-means clustering algorithms, principal and independent component analysis and association rules are examples of unsupervised machine learning.

- Reinforced learning: is a machine learning technique that aims to create autonomous agents capable of choosing actions to perform to achieve certain objectives through interaction with the environment in which they are immersed. It is one of the three main paradigms of machine learning, along with supervised and unsupervised learning. Unlike the other due ones, this paradigm deals with sequential decision problems, in which the action to be performed depends on the current state of the system and determines its future.

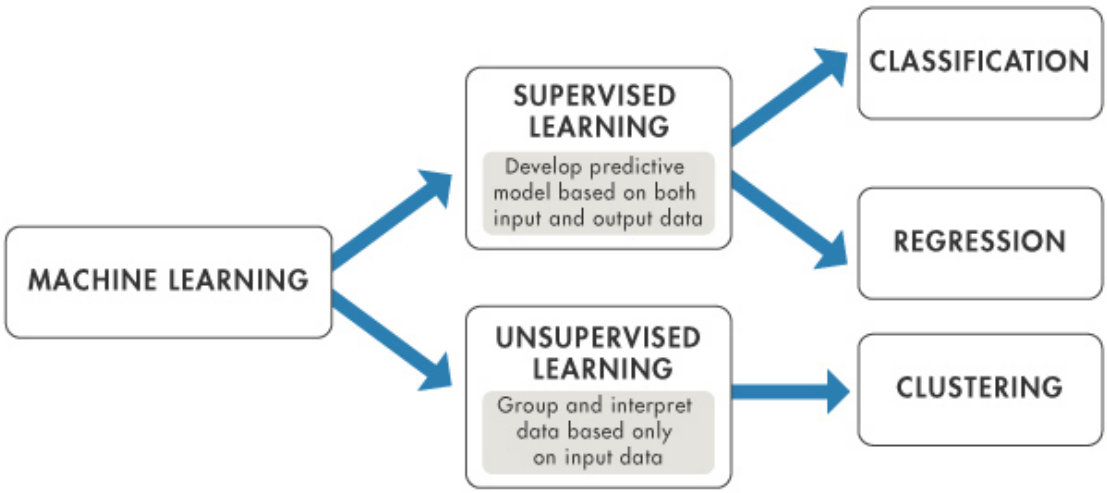


Figure 3.2: Machine learning paradigms: supervised vs unsupervised [94]

Supervised learning can be separated into two types of problems, called regression and classification (see Figure 3.2):

- Regression is used to understand the relationship between dependent and independent variables. It is commonly used to make predictions, such as for sales revenue for a given business. Examples of popular regression algorithm are: linear regression, logistical regression, and polynomial regression.
- Classification uses an algorithm to accurately assign test data into specific categories. It recognizes specific entities within the dataset and attempts to draw some conclusions on how those entities should be labeled or defined. Common classification algorithms are support vector machines (SVM), decision trees, k-nearest neighbor, and random forest.

For this thesis work we will focus exclusively on supervised learning and, specifically, on regression algorithms.

3.2 Tools for machine learning modeling

This second paragraph describes all the tools that are to be exploited in support of or following the training of machine learning algorithms; to name but a few that will be described later: metrics for evaluating the performance of predictive models, correlation coefficients for analyzing datasets and choosing the input variables to the models among the various candidates, machine learning regression models and training approacehs, and so on.

3.2.1 Machine learning algorithms

Multiple Linear Regression (MLR)

Linear Regression (LR) models are ML models which establish a linear relation between a dependent variable, typically called the target variable, and one or several independent variables, namely the features. The equation of a LR model can be generally written as reported in (3.1); where y is the target variable, x is the input feature belonging to a group of n features, a represents the model coefficient and ε the stochastic error.The function can also be expressed using normal notation, see (3.2).

$$y = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_n x_n + \varepsilon \tag{3.1}$$

$$y = X\beta + \varepsilon \tag{3.2}$$

The advantages of Multiple Linear regression are:

- Simple and fast implementation.
- Performs best on Linear Data.

The disadvantages of Multiple Linear regression include:

- Prone to underfitting.
- Sensitive to outliers.
- Assumes that data is independent.

Artificial Neural Networks (ANNs)

Artificial neural networks (ANNs), usually simply called neural networks (NNs) are computing systems inspired by the biological neural networks that constitute human brain. An ANN is based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain, can transmit a signal to other neurons.

There are many types of neural networks available or that might be in the development stage. They can be classified depending on their: structure, data flow, neurons used and their density, layers and their depth activation filters etc. The main 9 typologies are the following:

- Perceptron

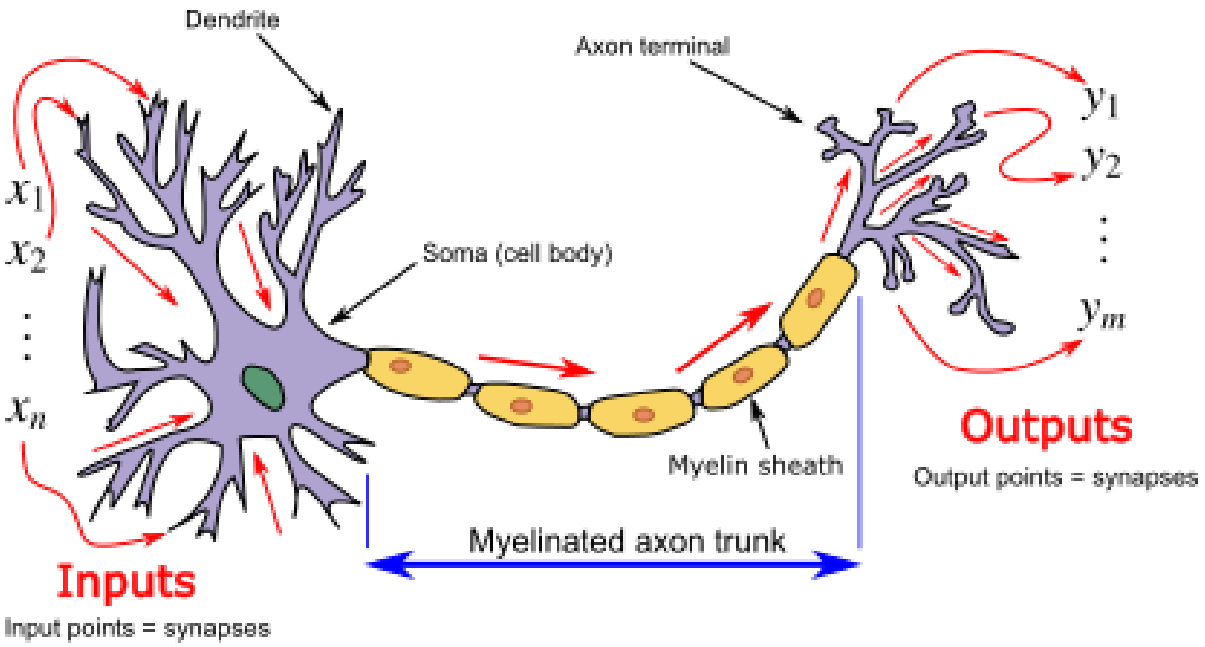


Figure 3.3: Neuron and myelinated axon, with signal flow from inputs at dendrites to outputs at axon terminals

- Feed Forward Neural Network
- Multilayer Perceptron
- Convolutional Neural Network (CNN)
- Radial Basis Functional Neural Network
- Recurrent Neural Network (RNN)
- LSTM – Long Short-Term Memory
- Sequence to Sequence Models
- Modular Neural Network

A Multi-layer Perceptron (MLP) is a fully connected class of feed-forward artificial neural network (ANN), which is s an artificial neural network wherein connections between the nodes do not form a cycle. As such, it is different from its descendant: recurrent neural networks. The feed-forward neural network was the first and simplest type of artificial neural network devised. In this network, the information moves in only one direction—forward—from the input nodes, through the hidden nodes (if any) and to the output nodes. There are no cycles or loops in the network. MLP actually is a supervised learning algorithm that learns a function $f(x) : R^n \longrightarrow R^o$ by training on a dataset, where m is the number of dimensions for input and o is the number of dimensions for output. Given a set of features and a target, it can learn a non-linear function approximator for either classification or regression. It is different from logistic regression, in that between the input and the output layer, there can be one or more non-linear layers, called hidden layers. Figure 3.4 shows a one hidden layer MLP with scalar output.

The leftmost layer, known as the input layer, consists of a set of neurons $\{x_i|x_0, x_1, ..., x_n\}$ representing the input features. Each neuron in the hidden layer transforms the values from the previous layer with a weighted linear summation $w_1x_1 + w_2x_2 + ... + w_nx_n$, followed by a non-linear activation function - like the hyperbolic tan function. The output layer receives the values from the last hidden layer and transforms them into output values. The advantages of Multi-layer Perceptron are:

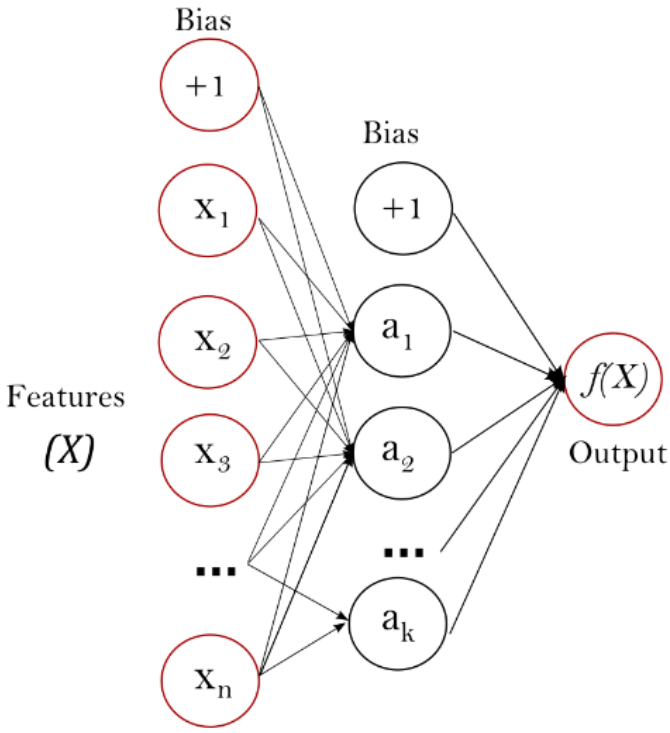


Figure 3.4: Multi-layer Perceptron layout with one hidden layer [81]

- It can be used to solve complex nonlinear problems.
- It handles large amounts of input data well.

The disadvantages of Multi-layer Perceptron (MLP) include:

- MLP with hidden layers have a non-convex loss function where there exists more than one local minimum. Therefore different random weight initialization can lead to different validation accuracy.
- MLP is sensitive to feature scaling.
- MLP requires tuning a number of hyperparameters such as the number of hidden neurons, layers, and iterations.

Focusing on the latter disadvantage, it is important to define what an hyperparameter is. A hyperparameter is a parameter whose value is used to control the learning process. By contrast, the values of other parameters (typically node weights) are derived via training. Hyperparameters can be classified as model hyperparameters, that cannot be inferred while fitting the machine to the training set because they refer to the model selection task, or algorithm hyperparameters, that in principle have no influence on the performance of the model but affect the speed and quality of the learning process. An example of a model hyperparameter is the topology and size of a neural network. Examples of algorithm hyperparameters are: activation functions, learning rate, batch size, hydden layer sizes and so on. In artificial neural networks, the activation function of a node defines the output of that node given an input or set of inputs. A standard integrated circuit can be seen as a digital network of activation functions that can be "on" or "off" , depending on input. Main activation functions commonly used in literature are enlisted below and shown also in figure 3.5.

- Logistic: $f(x) = \frac{1}{1+e^{-x}}$

- Relu: $f(x) = max(0, x)$
- Linear: $f(x) = x$
- Tanh: $f(x) = tanh(x)$

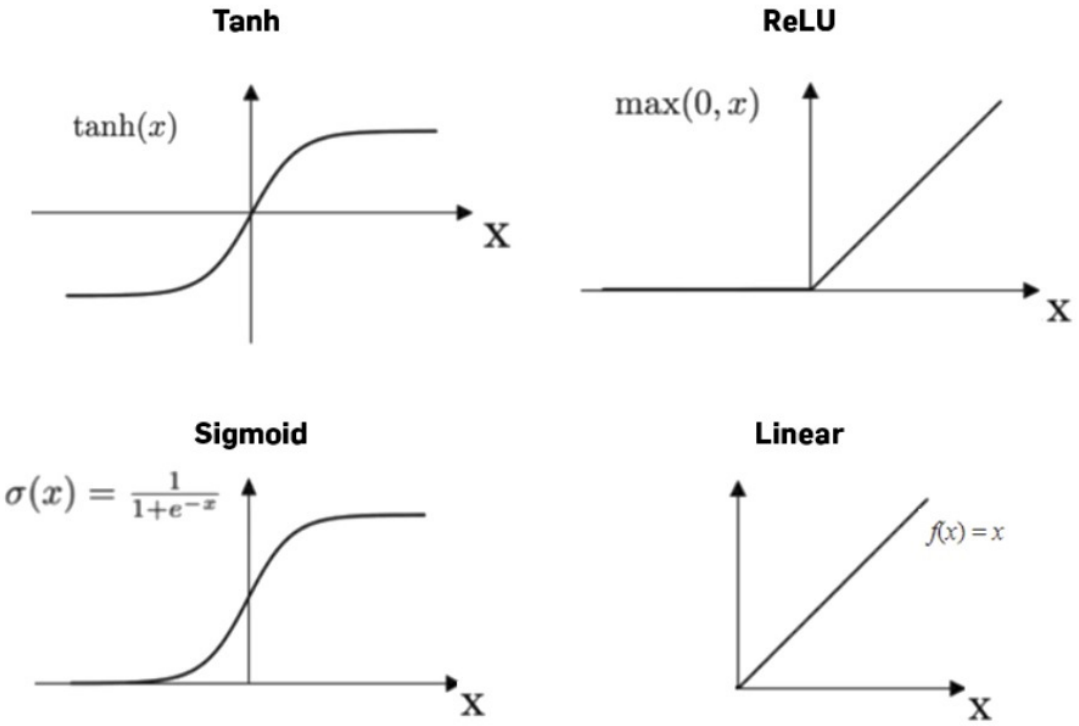


Figure 3.5: Typical activation functions used for ANN layers

The way a NN is trained can be explained introduction the concept of back-propagation. Back-propagation computes the gradient of a loss function with respect to the weights of the network for a single input–output example, and does so efficiently, computing the gradient one layer at a time, iterating backward from the last layer to avoid redundant calculations of intermediate terms in the chain rule; this can be derived through dynamic programming. The term back-propagation strictly refers only to the algorithm for computing the gradient, not how the gradient is used; however, the term is often used loosely to refer to the entire learning algorithm, including how the gradient is used, such as by stochastic gradient descent. The learning rate is a hyper parameter that controls how much to change the model in response to the estimated error each time the model weights are updated. Choosing the learning rate is challenging as a value too small may result in a long training process that could get stuck, whereas a value too large may result in learning a sub-optimal set of weights too fast or an unstable training process.

The number of neurons in the hidden layers (HLN) of an artificial neural network is an important architectural hyperparameter that can significantly impact the model's ability to learn and generalize to unseen data. The hidden layers of an ANN are responsible for learning and representing the underlying patterns and relationships in the input data. Traditionally, the number of neurons in the hidden layers of an ANN has been chosen heuristically, based on the size and complexity of the input data and the desired model capacity. However, these must be chosen, along with all other hyperparameters, during the hyperparameter tuning phase.

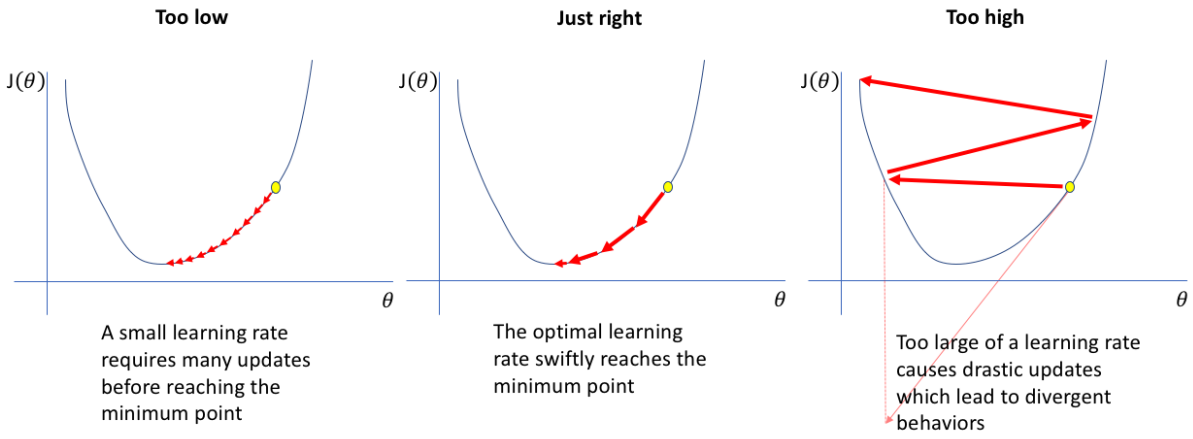


Figure 3.6: Learning rate effect on solution convergin speed and effectiveness

Training, validation and testing of machine learning models

The input data used to build the model are usually divided into multiple data sets. In particular, three data sets are commonly used in different stages of the creation of the model: training, validation, and test sets. The model is initially fit on a training data set, which is a set of examples used to fit the parameters (e.g. weights of connections between neurons in artificial neural networks) of the model. The model is trained on the training data set using a supervised learning method, for example using optimization methods such as gradient descent or stochastic gradient descent. In practice, the training data set often consists of pairs of an input vector (or scalar) and the corresponding output vector (or scalar), where the answer key is commonly denoted as the target (or label). The current model is run with the training data set and produces a result, which is then compared with the target, for each input vector in the training data set. Based on the result of the comparison and the specific learning algorithm being used, the parameters of the model are adjusted. The model fitting can include both variable selection and parameter estimation. Successively, the fitted model is used to predict the responses for the observations in a second data set called the validation data set. The validation data set provides an unbiased evaluation of a model fit on the training data set while tuning the model's hyperparameters (e.g. the number of hidden units—layers and layer widths—in a neural network). Validation datasets can be used for regularization by early stopping (stopping training when the error on the validation data set increases, as this is a sign of over-fitting to the training data set). This simple procedure is complicated in practice by the fact that the validation dataset's error may fluctuate during training, producing multiple local minima. This complication has led to the creation of many ad-hoc rules for deciding when over-fitting has truly begun. Finally, the test data set is a data set used to provide an unbiased evaluation of a final model fit on the training data set. If the data in the test data set has never been used in training (for example in cross-validation), the test data set is also called a holdout data set. The term "validation set" is sometimes used instead of "test set" in some literature (e.g., if the original data set was partitioned into only two subsets, the test set might be referred to as the validation set). Deciding the sizes and strategies for data set division in training, test and validation sets is very dependent on the problem and data available. Consequently, to correctly perform the training process, the available dataset must be divided into three parts as follows:

- the training set (70%): shown to the algorithm to perform the learning
- the validation set (10%): used to evaluate the performance of the model at the end

- the test set (20%): used at the end of the process to evaluate the prediction performance of the trained model. of each iteration of the learning process

Performance evaluation during the learning process is very important to ensure that the process goes smoothly. The main risk is that the model focuses too much on the data it is shown, failing to predict instants with different characteristics from those in the training dataset, and losing generality. A correct learning process, consequently, must exhibit consistent variations in the loss function (i.e., the chosen error metric which, in this case, is the mean squared error) along the various Epochs (iterations). From the moment that a continued increase in validation loss occurs with a continued reduction in training loss, the model will have lost its effectiveness on data not previously seen (overfitting).

3.2.2 Features selection methods

Feature Selection and Engineering is a key step for a machine learning model development project, since it allows to choosing of the optimal subset of input features starting from the raw data database and permits increasing the data-driven model interpretability and generalization, adding the previous knowledge to increase the transparency of the input-output relation, and that's why it is strongly recommended to follow a systematic method of choosing the starting subset. Feature selection methods are intended to reduce the number of input variables to those that are believed to be most useful to a model to predict the target variable. One way to think about feature selection methods is in terms of supervised and unsupervised methods. An important distinction to be made in feature selection is that of supervised and unsupervised methods. When the outcome is ignored during the elimination of predictors, the technique is unsupervised. The difference has to do with whether features are selected based on the target variable or not. Unsupervised feature selection techniques ignore the target variable, such as methods that remove redundant variables using correlation. Supervised feature selection techniques use the target variable, such as methods that remove irrelevant variables. Another way to consider the mechanism used to select features which may be divided into wrapper and filter methods. These methods are almost always supervised and are evaluated based on the performance of a resulting model on a hold-out dataset. Wrapper feature selection methods create many models with different subsets of input features and select those features that result in the best-performing model according to a performance metric. These methods are unconcerned with the variable types, although they can be computationally expensive. RFE is a good example of a wrapper feature selection method [94]. All these cases are described in Figure 3.7.

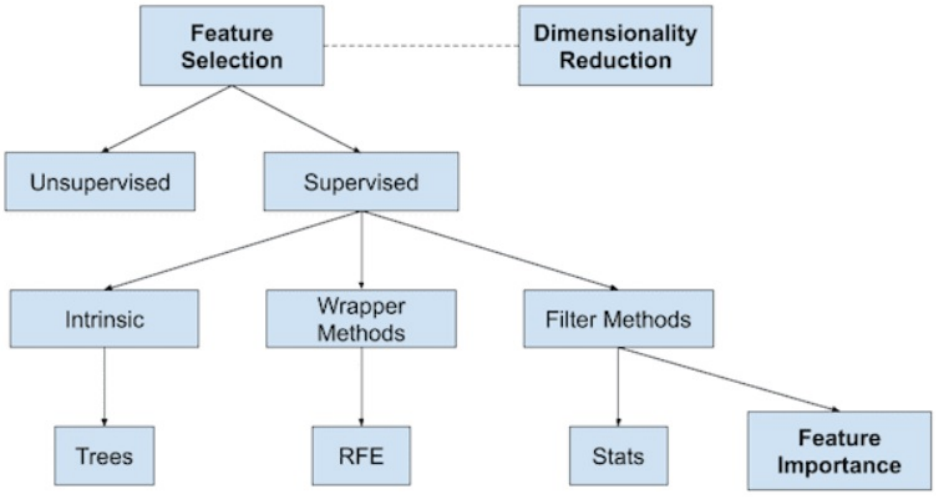


Figure 3.7: Features selection methods [94]

Figure 3.8 shows which are the best tools to carry on a filter method when the configuration of Input/Output is : Numerical Input -Numerical Output. This is a regression predictive modeling problem with numerical input variables. The most common techniques are to use a correlation coefficient, such as Pearson's for a linear correlation, or rank-based methods for a nonlinear correlation, such as Spearman correlation index. Exactly these two will be used in this thesis work and will be described in the following section.

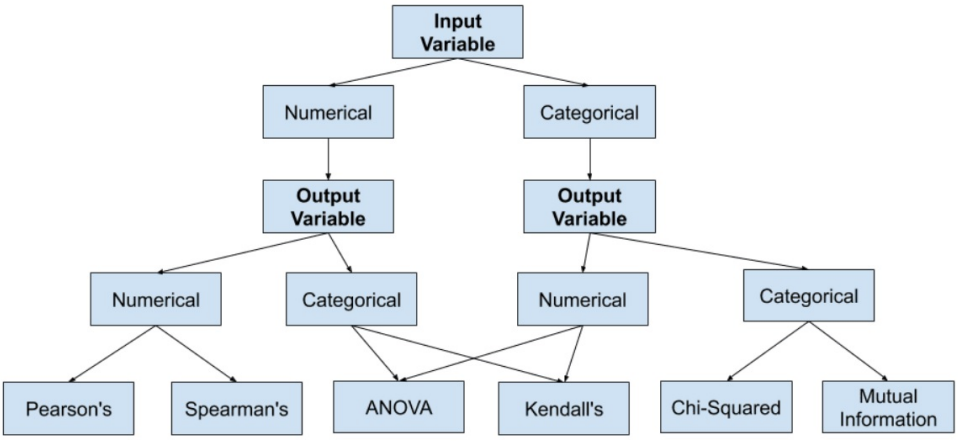


Figure 3.8: Filter methods and corresponing tools and correlation indexes according to different pairs of input and output datatypes [94]

3.2.3 Correlation coefficients

Several different types of coefficients assessing the correlation existing between two populations of data are present in literature: for this study the authors decided to take advantage of Pearson and Spearman correlation coefficients.

- Pearson product-moment correlation coefficient is a measure of the linear correlation between two set of data and is given by the ratio between the covariance of two variables (X, Y) and their standard deviation product. The Pearson coefficient can only assess a linear correlation, ignoring all other types (quadratic correlation, etc...).

$$r\left(X_i, Y_i\right) = \frac{cov\left(X, Y\right)}{\sigma_X \cdot \sigma_Y} = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}} \tag{3.3}$$

- Spearman's rank correlation coefficient is a measure of rank correlation, that is the statistical dependence between the rankings of two variables and is given by the Pearson correlation coefficient of the two ranks. It assesses how well the relationship between two variables can be described using a monotonic function (whether linear or not) and it's also less sensitive to strong outliers than the Pearson correlation.

$$\rho\left(X_i, Y_i\right) = \frac{cov\left(R(X), R(Y)\right)}{\sigma_{R(X)} \cdot \sigma_{R(Y)}} \tag{3.4}$$

3.2.4 Dataset scaling

Standardization of a dataset is a common requirement for many machine learning estimators: they might perform badly if the individual features do not more or less look like standard normally distributed data (e.g. Gaussian with 0 mean and unit variance).

$$x_{scal,std} = (x - u)/\sigma \tag{3.5}$$

Sometimes an input variable may have outlier values. These are values on the edge of the distribution that may have a low probability of occurrence. Outliers can skew a probability distribution and make data scaling using standardization difficult as the calculated mean and standard deviation will be skewed by the presence of the outliers. One approach to standardizing input variables in the presence of outliers is to ignore the outliers from the calculation of the mean and standard deviation, then use the calculated values to scale the variable. This is called robust standardization or robust data scaling. This can be achieved by calculating the median (50th percentile) and the 25th and 75th percentiles. The values of each variable then have their median subtracted and are divided by the interquartile range (IQR) which is the difference between the 75th and 25th percentiles.The robust scaler removes the median and scales the data according to the quartile range (defaults to IQR: Interquartile Range). The IQR is the range between the 1st quartile (25th quartile) and the 3rd quartile (75th quartile).

$$x_{scal,rob} = (x - u)/(p_{75} \breve{} p_{25}) \tag{3.6}$$

The resulting variable has a zero mean and median and a standard deviation of 1, although not skewed by outliers and the outliers are still present with the same relative relationships to other values. Both of these methods were exploited via the Scikit-Learn dedicated library [81].

3.2.5 Cross-validation and one-hot encoding

Cross-validation is a statistical technique that can be used in the presence of a huge observed sample size. In particular, so-called k-fold cross-validation consists of dividing the total data set into k parts of equal samples and, at each step, the k part of the data set comes to be the validation part, while the remaining part always constitutes the training set. Thus, one trains the model for each of the k parts, thus avoiding problems of overfitting, but also of asymmetrical (and thus biased) sampling of the observed sample, which is typical of splitting the data into only two parts (i.e. training/validation). In other words, one divides the observed sample into groups of equal numerosity, iteratively excludes one group at a time, and tries to predict it with the non-excluded groups, in order to verify the goodness-of-fit of the prediction model used. One-hot encoding is a method of data presentation for which the features are encoded using a one-hot (aka 'one-of-K' or 'dummy') encoding scheme. This creates a binary column for each category and returns a sparse matrix or dense array. A classic example concerns the days of the week or the various months, instead of using a single vector ranging from 1 to n, n vectors of 0 or 1 are used. Figures 3.9 explains better what is one-hot encoding.

3.2.6 Prediction quality evaluation indices

To evaluate the predictive accuracy of the developed model, several performance indicators are introduced and used for this work, according to following equations, where y_i is the real value, $\bar{y}_i$ the mean value, y^* the predicted value and N represents the dataset size.

- Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{100\%}{N} \sum_{i=0}^{N-1} \frac{y_i - \hat{y}_i}{y_i}. \tag{3.7}$$

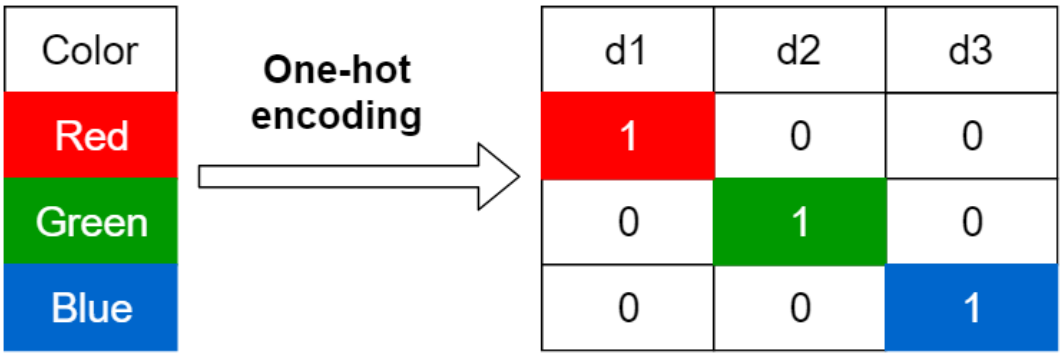


Figure 3.9: One-hot encoding explanation [95]

- Mean Absolute Error (MAE)

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i^* - y_i| \tag{3.8}$$

- Mean Squared Error (MSE)

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i^* - y_i)^2 \tag{3.9}$$

- Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i^* - y_i)^2} \tag{3.10}$$

- R^2 or Coefficient of Determination (CD)

$$R^2 = 1 - \frac{\frac{1}{N}\sum_{i=1}^{N}(y_i^* - y_i)^2}{\frac{1}{N}\sum_{i=1}^{N}(\bar{y_i} - y_i)^2} \tag{3.11}$$

The MAPE index is massively used in literature for models' performance evaluation since it's easily interpretative and gives a rapid overview of the prediction errors in terms of percentage values. MAE is the average absolute error between actual and predicted values: the closer MAE is to 0, the more accurate the model is. But MAE is returned on the same scale as the target you are predicting for and therefore there isn't a general rule for what a good score is. How good your score is can only be evaluated within your dataset. RMSE is obtained from the MSE value, and it gives immediately the measure of the difference between the values of two populations of data, that is the predicted and real values groups. The coefficient of determination R2 is a statistical parameter and provides a measure of how well-observed outcomes are replicated by the model, assuming values from 0 to 1.

3.3 Energy monitoring and targeting

Energy monitoring is primarily a management technique that enables industries to control energy consumption accurately. Monitoring is an approach in energy management to eliminate waste, reduce current level of energy use and improve the existing operating procedures. Monitoring and targeting (M&T) is an activity, which uses information on energy consumption as a basis for the control and management of energy use and is based on the management principle that effective management requires accurate measurement. It essentially combines the principles of energy use and statistics. Monitoring is essentially aimed at preserving an established pattern. Target setting is the identification of energy consumption level, which is desirable as a management objective to work towards. M&T techniques are based on some key steps that create a continuous feedback loop, thereby enhancing the control of energy utilization.

- Measurement and data gathering: the initial step involves the continuous metering of key quantities and the gathering of data from various meters.
- Baseline energy modeling: the collected data are used to define the baseline. This is essentially a mathematical model that represents the energy demand of the system based on the identified energy drivers, that is, those parameters that influence its performance. The methods used can vary, including machine learning-based ones. The resulting model will then be used as an indicator of the system energy demand in "standard" conditions.
- Monitor variations: the subsequent step involves monitoring the discrepancy between the predicted energy consumption and the actual consumption as recorded by meters. One commonly utilized tool for this purpose is the CuSuM CUmulative SUM), which stands for the cumulative sum of differences. CuSuM charts are often used when the detection of small changes in a process parameter is important [96].
- Causes identification and targets definition: the CUSUM graph is exploited to determine the root causes of deviations in energy consumption. Such deviations may be due to changes in faulty system behaviors, process modifications, alterations in external conditions, and so on. Upon establishment of the baseline and identification of factors affecting energy consumption, it becomes possible to set meaningful targets for future consumption.

3.3.1 CuSuM control chart

A CuSuM chart is a plot of cumulative sequential differences between each data point and process average over time. CuSuM is a technique to see through random scatter and to detect changes in the pattern in monitoring the energy consumption and helps target setting. The CuSum method consists of calculating the differences between the real consumption and the predicted one and using them to generate the control chart. The following equation shows how the CuSum is calculated for the dataset:

$$S_j = S_{j-1} + \varepsilon_j = S_{j-1} + (C_j - C_{std,j}) \tag{3.12}$$

where S_j is the cumulated sum of the j-th time-step, C_j is the current consumption value, $C_{std,j}$ is the standard consumption calculated for the current time-step with a model and ε_j is the current residual value. Therefore, a positive S_j will imply a higher energy consumption than the standard one and a negative S_j will imply a lower energy consumption than the expected one. The CuSum control chart is a memory control chart and can quickly detect small and moderate shifts in the considered process.

Chapter 4

Case studies analysis

This chapter presents the real datasets used for this thesis and the results of their analysis. The material provided by the DSO relates to the operation of seven CGSs for at least one year. The various CGSs are located in central Italy and serve regions belonging to predominantly D and E climate zones. The cluster of cabins supplied includes different sizes in terms of gas volume flow and pressure differences. In addition, the actual consumption of the preheating plant is also reported for two of the CGSs. From now on, the plants will be labeled with numbers from 1 to 7 to distinguish them for the techno-economic analysis (chapter 5). Plants 1 and 2 will be renamed A and B for energy monitoring analysis using machine learning algorithms (chapter 6).

4.1 Schematic description of a CGS thermal plant

Figure 4.1 shows a typical layout of a CGS plant. The system includes the High Pressure (HP) inlet, two redundant lines with gas Filters (F), Preheaters (PH), gas expansion Valves (V), and the stations for Fiscal Measurement (FM) and Odorant (OD) injection before the gas is fed back into the low pressure (LP). The control logic of the installed system does not provide for an inverter-controlled flow rate, as assumed in the theoretical study, but a constant value of the preheating water flow rate regardless of the gas conditions at the CGS inlet. This affects preheating efficiency, as the minimum flow rate of water required for preheating is never supplied, and the two pumps (P1 and P2) always run at constant revolutions and process the same water flow rate. Boilers generally operate alternately: the one or those, with three or more boiler layouts, that do not operate, remain in standby mode and consume an almost constant amount of gas for the pilot flames, estimated by the DSO to be about 0.25 Smc/h.

Although this plant represents a standard for the many CGSs scattered across Italian territory, some plant details may vary and consequently impact the actual energy consumption to manage to preheat. Table 4.1 shows a list of possible monitored variables for the operating of the CGSs. Several temperatures could be measured, such as the environmental air temperature (T_{env}), the gas outlet temperature (T_{gas}), the gas stream outlet set point (T_{SP}) and the preheating water temperature (T_{wat}) with its set point if a thermal control system is working in that specific plant ($T_{wat,SP}$). Even though the gas inlet conditions are crucial to calculate the preheating requirements, it is less common to measure the natural gas inlet temperature. To avoid this issue, several academic studies correlate the environmental air temperature with the soil depth to obtain the gas inlet temperature. For this work, a data-driven correlation suggested by the partner DSO is used in the thermodynamic equations, as already described in chapter 2. On the other hand, inlet pressure value (P_{in}) is usually monitored by the DSO and, along whit this, the outlet pressure (P_{out}) is given as well. The main flow rates that are monitored are the natural gas flow rate

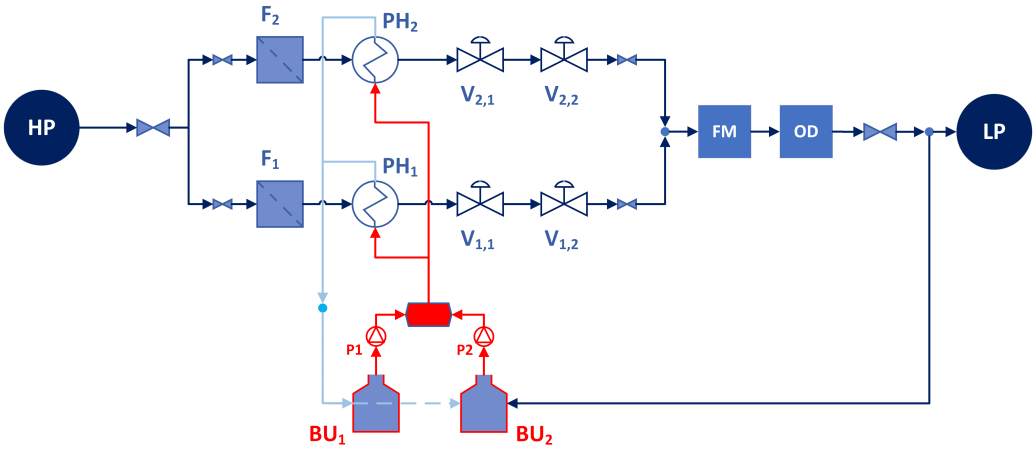


Figure 4.1: Italian CGS typical layout

passing through the plant (Q_{gas}) and the natural gas flow rate feeding the boiler units for preheating requirements (Q_{BU}). Both of them are calculated in terms of standard cubic meters per hour, since the fiscal meter is designed to compute the energy passing through the CGS.

Table 4.1: List of possible signals acquired in a standard CGS

ID	Name	Unit
P_{in}	Gas relative inlet pressure	barg
P_{out}	Gas relative outlet pressure	barg
T_{env}	Environmental air temperature	°C
T_{gas}	Gas outlet temperature	°C
$T_{gas,in}$	Gas inlet temperature	°C
T_{SP}	Gas outlet temperature SET POINT	°C
T_{wat}	Preheating water temperature	°C
$T_{wat,SP}$	Preheating water temperature SET POINT	°C
G	Solar irradiance	W/m^2
Q_{gas}	NG standard volumetric flow rate (gas flow)	Smc/h
C_{min}	Boiler on-off sensor	min/h
Q_{BU}	NG standard volumetric flow rate (boiler units consumption)	Smc/h

4.2 Case studies: real CGS datasets

Figure 4.2 shows the daily demand curve each CGS (rows) and each season (columns). Several conclusions can be drawn from the analysis of this graph: there are at least 2 peaks approximately located in the same time slots and which are shared by all the systems analyzed, these relate to the morning (from about 7 to 9) and the evening (roughly from 18/19 in the afternoon on-wards until 22). The motivation behind these peaks is that, despite the variability of users downstream for these cabins, the main consumption is always found in the morning (when everyone wakes up and uses the domestic hot water for washing

and preparing breakfast) and in the evening (returning from work and for dinner). Some plants, on the other hand, may also have a third peak, around noon; the latter may be due to an agglomeration of predominantly domestic users which also expects a peak for lunchtime, but which tends to be lower than the two extreme peaks. All the previous considerations acquire greater validity in the winter seasons when consumption is very high and the dependence on the external temperature is marked and evident. For the summer seasons, on the other hand, the demand for natural gas can also drop considerably, except for plants that serve various industrial users downstream, which have load profiles that vary less with the seasons and during the day itself.

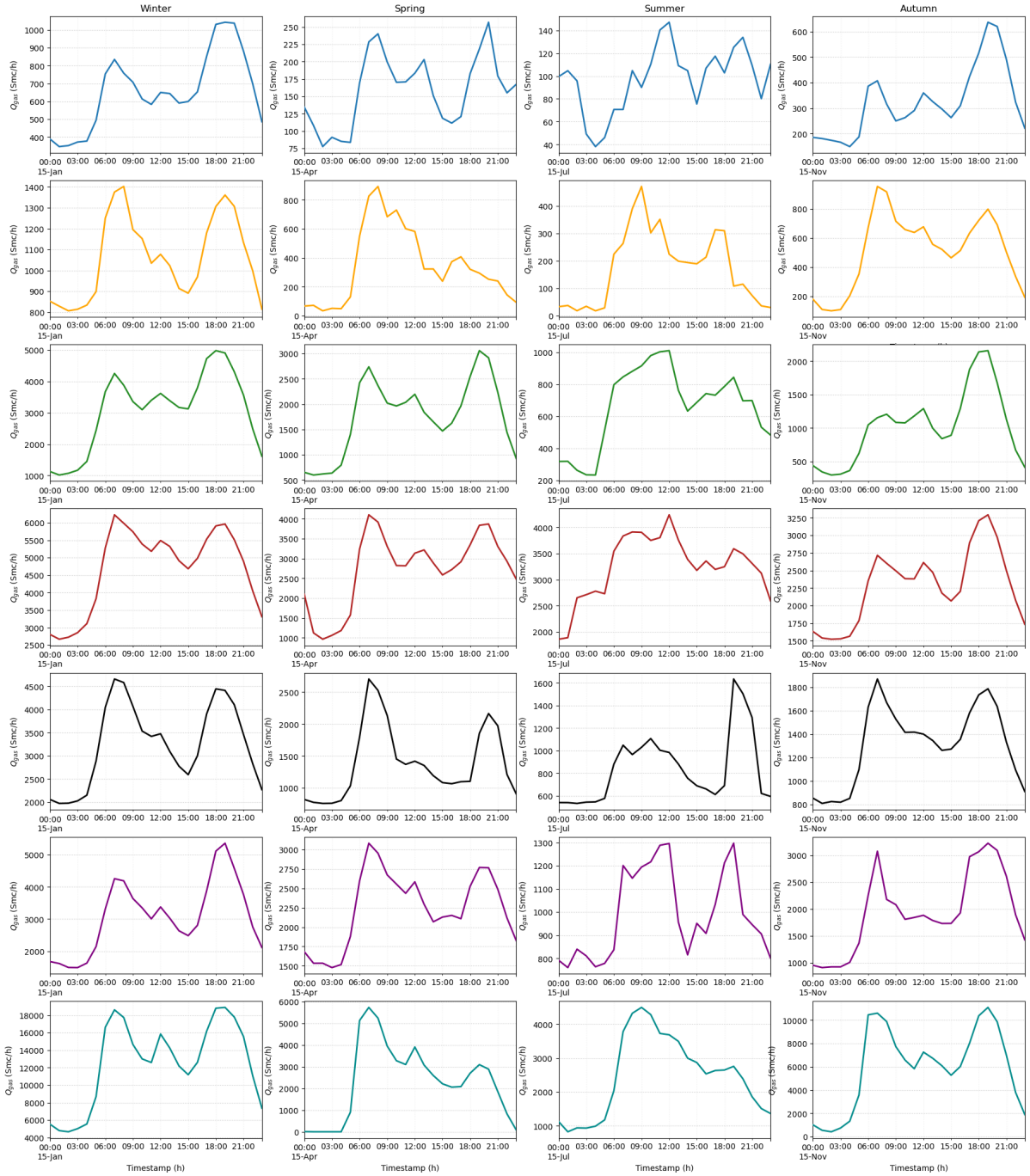


Figure 4.2: CGS volumetric flow rates: plant (rows) vs season (columns)

Figure 4.3 shows the yearly variation of the natural gas flow rate for each plant. Despite small differences among the considered plants, there are two main thermal seasons: the winter one which starts between October and November and ends between March and April, and the summer one, which lasts for

the remaining period and is characterized by mean values that could also be 1/4 of the mean values in the winter season. During the summer season, there is no dependence between the external temperature and NG consumption, because the gas heating systems are switched off from a certain period of spring onwards and switched on only in late autumn. This goes back to what was described in the section on climate zones on Italian scenario in Chapter 1.

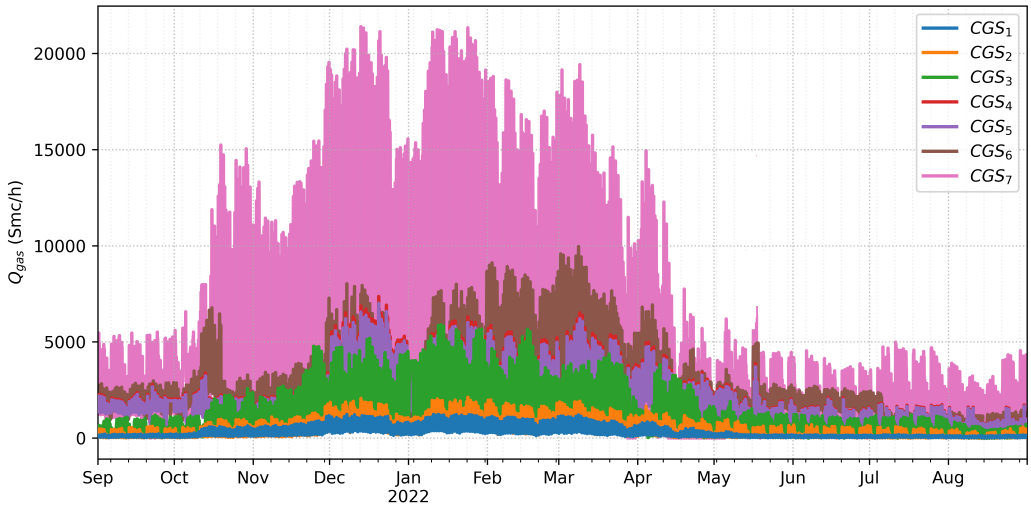


Figure 4.3: CGS volumetric flow rates

Table 4.2 reports the mean values of the main parameters describing the entire annual operating conditions of each CGS, such as the natural gas flow rate that passes through the plant Q_{gas}, the gas inlet pressure P_{in}, and the set point of the gas outlet pressure P_{out}. Cabin sizes vary from a minimum of a few hundred cubic meters per hour for the CGS_1 (300 Smc/h) and CGS_2 (500 Smc/h) to a maximum of several tens of thousands of cubic meters per hour of gas for the CGS_7 (5700 Smc/h).

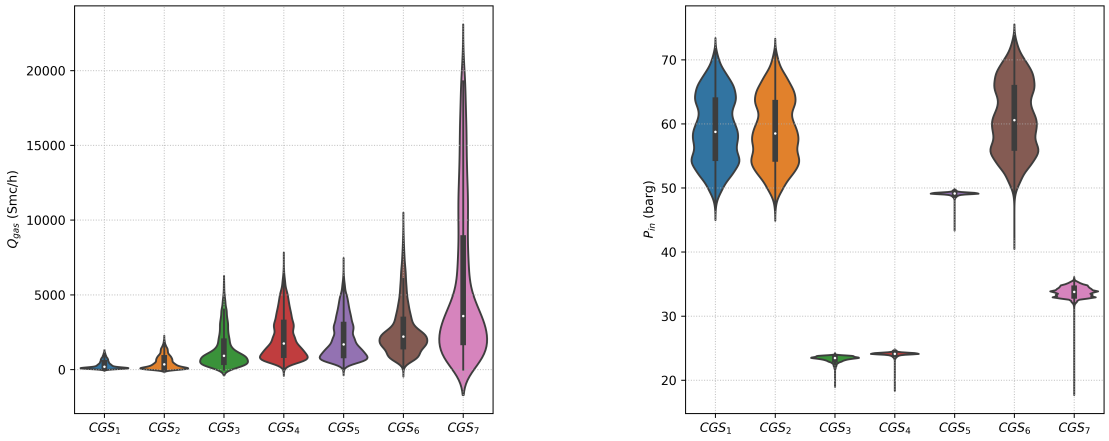
Table 4.2: CGS operating conditions: mean values

CGS_{ID}	$Q_{gas}[Smc/h]$	$P_{in}[barg]$	$P_{out}[barg]$
CGS_1	298.269	59.124	3.149
CGS_2	493.352	58.876	3.336
CGS_3	1334.686	23.364	3.999
CGS_4	2170.937	24.112	3.985
CGS_5	2086.390	49.114	3.980
CGS_6	2644.297	60.905	3.161
CGS_7	5712.421	33.774	4.117

The pressure values are more similar as there are few options in terms of the pressure drops for a CGS. The maximum pressure drop for the NG is experienced in the CGS_6 (about 57 barg), while the minimum pressure reduction is done in the CGS_4 (About 20 barg). This variation in the maximum pressure drop from one plant to another depends on various factors, including the position of the CGS in relation to the main pipeline, the volumes of gas processed during the year, and the plant layout available, such as the

capacity of pressure reduction monitors and pre-heating and filtering systems. The outlet pressure, on the other hand, is between 3 and 4 barg for all cabins, which perfectly reflects the positioning of these plants in the Italian gas infrastructure: the medium-low pressure networks immediately downstream of the cabins themselves generally operate at a few bars of pressure. This is due to the fact that large volumes of gas are still being transported, after which the flows will split and the last pressure variation (from about 3/4 barg to a few tens of millibars) will be met in small stations, called PRS (Pressure Reducing Stations), which do not need to preheat the gas given the small pressure jump they have to make and the negligible Joule-Thomson effect that follows.

Figures 4.4a and 4.4b are violin plots that help to understand the differences among the sample of plants that are considered for this thesis work. Our cluster of datasets that can be used for the work shows three more or less distinct groups: a first group with very small sizes in terms of processed gas flow rates and also quite marked variations between maximum and minimum (see plants 1 and 2), then there is a second group (2 to 6) and finally the last group (cabin 7). Some plants show less elongated and more "bellied" flow rate distributions signifying greater constancy of values, while others show more elongated shapes signifying greater variability of values. The inlet pressure is highly variable for plants 1,2 and 6, while for cabins 3,4,5 and 7 the fluctuations are greatly reduced, averaging around 25, 50, and 35 barg, respectively.

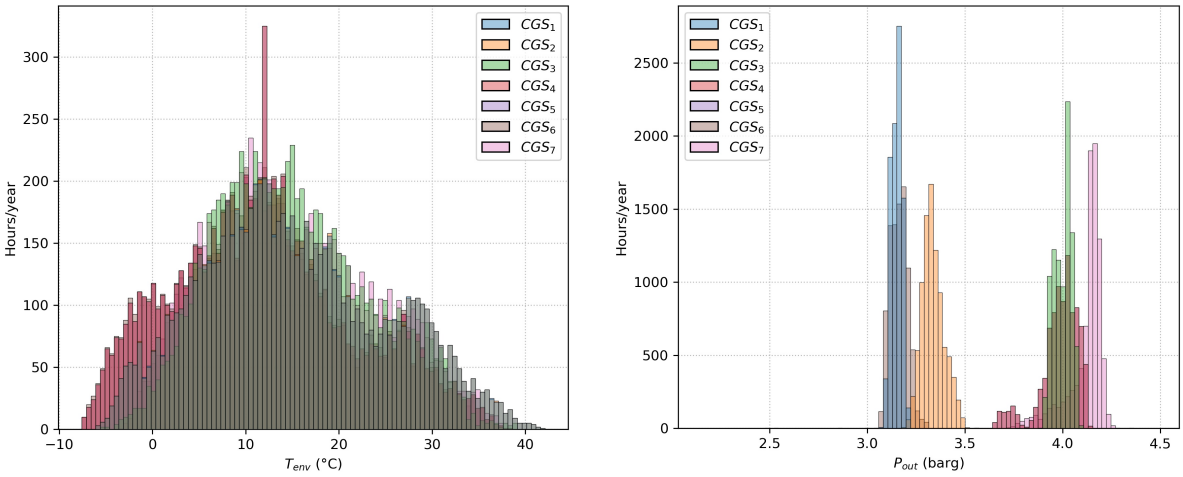


(a) (b)

Figure 4.4: CGS violin plots: volumetric flow rate (a) versus inlet pressure (b)

Figure 4.5a shows the statistical distribution of the outdoor air temperature for all the plants; since all the case studies belong to the same region and climatic zone, the distribution of temperatures is approximately the same in all seven stations. The environmental air temperature will vary between about -10°C to more than 37/38 °C with peaks of 40 °C. The most common values lie between 10°C and 20°C; in fact, the mean temperatures of the one-year datasets vary between 11°C and 18°C. The second figure of the plot, 4.5b, shows the distribution of the outlet pressures for all the plants; some systems accept larger fluctuations around the calibration value of pressure monitors, such as CGS_2 and CGS_4, and systems with a smaller distribution of output pressure values. Note that these are not the set point values, but the actual feedback values of the system, which means that they actually represent the gas outlet conditions of the various plants. Figure 4.6 shows the scatter plots of gas flow rate versus outdoor air temperature for all seven plants considered; in addition, data points belonging to the "summer" heating season (in red) and the "winter" climatic season are highlighted for all plots. The trend that emerges is extremely

clear: depending on the type of plant and its plant configuration and the type of downstream utilities, there are slightly different trends, but it is an inversely proportional trend to the gas consumed by the downstream users Qgas as a function of environmental temperature Tenv, and thus of seasonality. The air temperature reaches values as low as 0°C for the summer season, it runs from mid-March to mid-October. This is because the second part of March still experienced very low temperatures for the year and the geographical location considered. The table 4.3 shows the measuremnt position of all sensors and their sampling.



(a) CGS environmental air temperature distribution

(b) CGS outlet pressure distribution

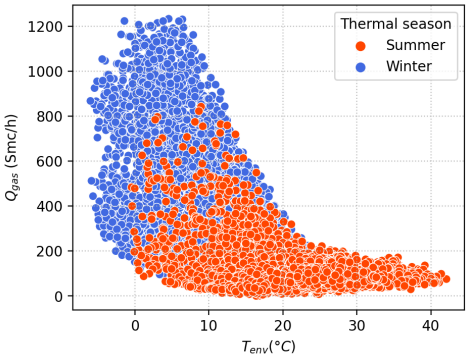
Figure 4.5: CGS outlet pressure distribution

4.2.1 Data set granularity

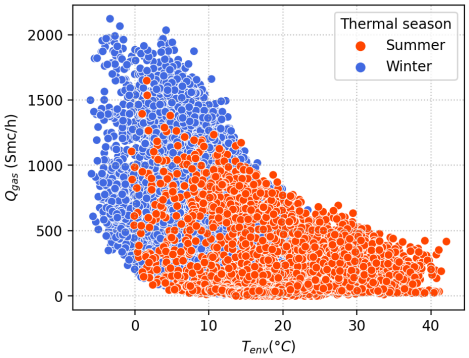
The proposed data set consists of the behavior of 7 plants for a period from the end of 2020 to the end of 2022. Data for raw measurements, such as flow rates of passing gas, pressures, and temperatures, are sampled every 15 minutes. In addition, several weather stations near the plants provide ambient temperature and irradiation levels with the same sampling. For this work purpose, the data will be used with hourly granularity, as the gas consumption of the thermal power plant is provided with a one-hour time step. The values will be averaged over the hour, reducing the data set of each plant from about 35000 data points to the canonical 8760 data points.

Table 4.3: Sensors location and dataset granularity

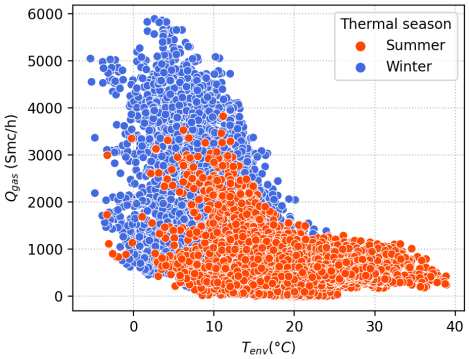
Symbol	Unit	Sensor location	Sampling
Q_{gas}	Smc/h	CGS	15 minutes
Q_{BU}	Smc/h	CGS (thermal plant)	1 hour
P_{in}	barg	CGS	15 minutes
P_{out}	barg	CGS	15 minutes
T_{gas}	°C	CGS	15 minutes
T_{env}	°C	Weather station	15 minutes
G	W/m^2	Weather station	15 minutes



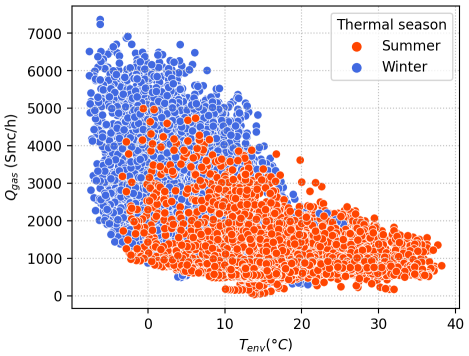
(a) CGS_1



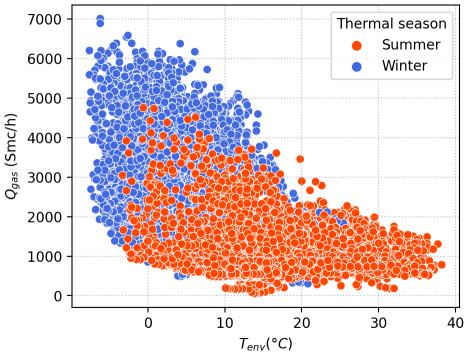
(b) CGS_2



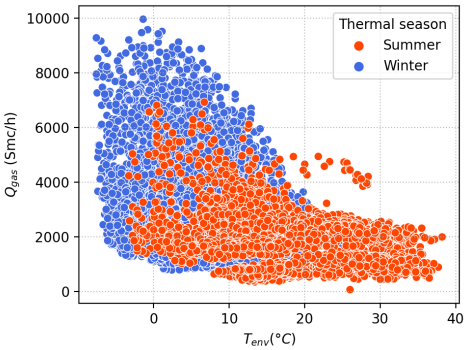
(c) CGS_3



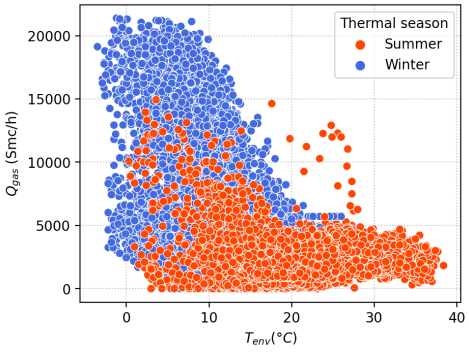
(d) CGS_4



(e) CGS_5



(f) CGS_6

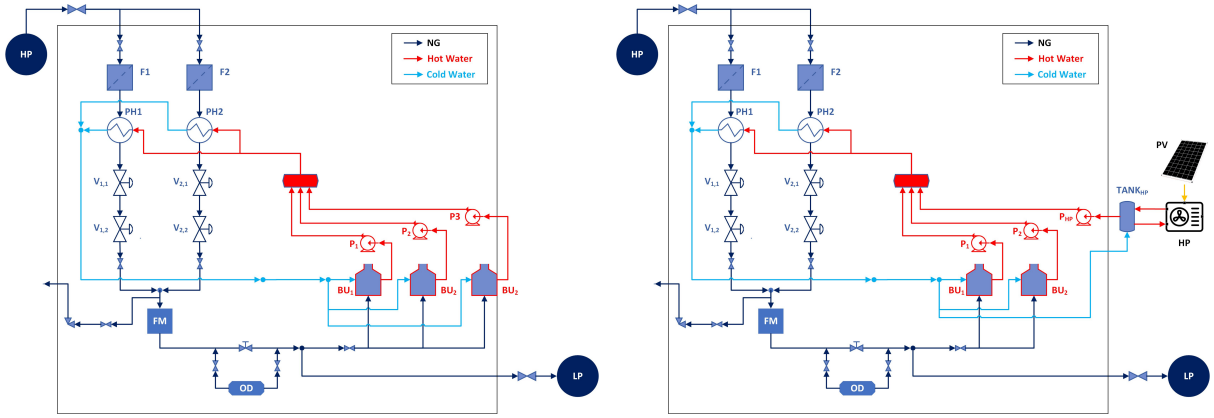


(g) CGS_7

Figure 4.6: NG volumetric flow rates vs environmental air temperatures for the two different thermal season (summer in orange and winter in blue)

4.3 Preheating systems dataset analysis

In this section, the datasets for the first two plants are analyzed, which also have data on their respective heating plants and thus heating consumption and all related parameters whenever possible. The installations involved are cabins CGS_1 and CGS_2, which were renamed CGS_A and CGS_B for the data-driven study and will be referred to as such from now on until the end of chapter two. Figure 4.7 shows the plant layouts of the two plants: plant A in Figure 4.7a and plant B in Figure 4.7b. The first layout is standard, while the second is an experimental layout, which was implemented by the DSO to study the possibility of accessing energy efficiency certificates. The main substantive difference is the extra boiler present in the first plant compared to the second plant, which instead has a system using a compression heat pump connected to an inertial tank; however, this heat pump was only active for a small portion of the time that is available in the datasets. In contrast, as far as all other devices are concerned, the two systems are perfectly equal.



(a) (b)

Figure 4.7: Plant specific layouts for CGS_A (a) and CGS_B (b), including thermal plant

Figure 4.9 and Figure 4.8 show the yearly trend of the NG flow rate and the pre-heating gas consumption, both in terms of Smc/h. The time period shown is the same chosen for the subsequent analysis and creation of the machine learning models: one year for training the models, specifically from April 2021 to April 2022, with the consecutive period used for testing. The preheating output is much higher in the winter season and tends to go to a very low steady state value in the summer season, as it is affected by the effect of the temperature increase, which has a twofold impact on preheating consumption: drastic reduction of the gas flow consumed by the utilities and higher processed gas temperatures.The peak hourly gas consumption reached by the cabins is respectively 4/5 for cabin A and 7/8 for cabin B. Moreover, the two cabins are located not many kilometres apart, so the temperature trends are practically the same throughout the year for both plants. This is an additional consideration that causes the consumption curves to be practically overlapping, barring errors and size-dependent scaling. Figure 4.10 displays the trend in gas temperature at the outlet of plants A and B. The outlet gas temperature of the former is much more variable and follows less the seasonality of the natural gas flow during the year. The second more closely follows the demand curve. This is because the first plant does not have a thermal power plant control system and the preheating water temperature set points are not variable; the second plant has feedback on the operating temperatures of the plant and this allows better utilization of the energy produced by the boilers. This can also be seen in the next picture, which shows the trends of the preheating water, i.e. the delivery of water to the pre-heater.

Figure 4.11 shows the trend of the pre-heating water temperature of the two plants A and B: the

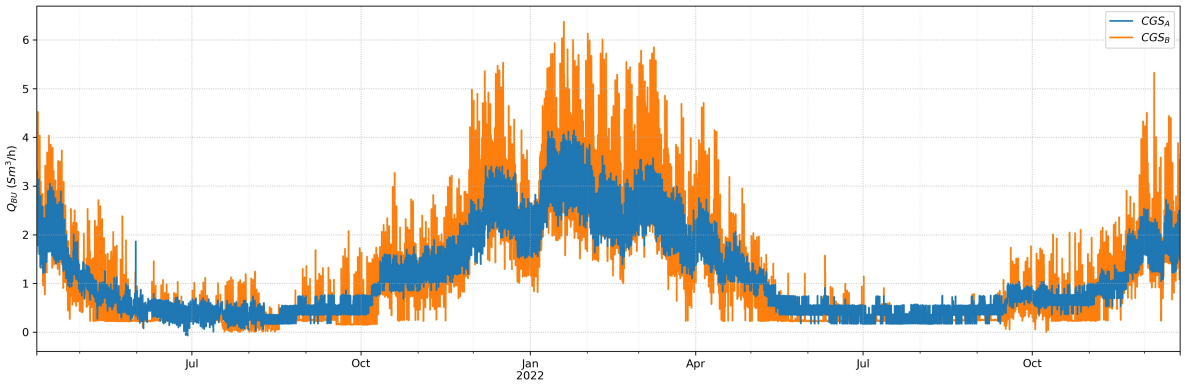


Figure 4.8: CGS_A vs CGS_B: natural gas consumption for preheating

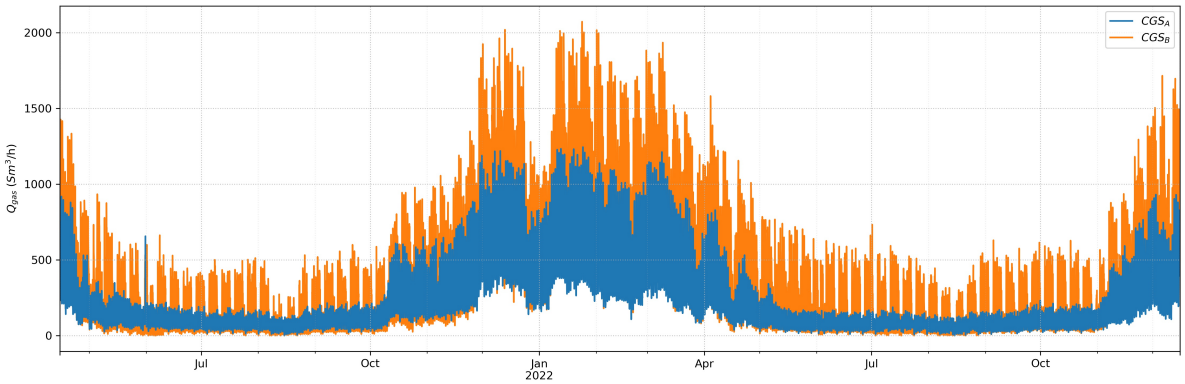


Figure 4.9: CGS_A vs CGS_B: gas volumetric flow rate

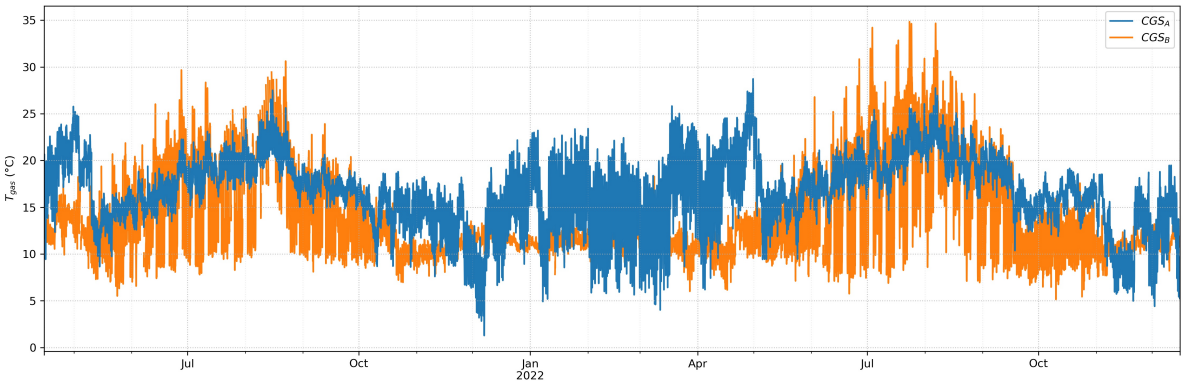


Figure 4.10: CGS_A vs CGS_B: outlet gas temperature

temperature of the first is kept constant as there is no control system according to the operating conditions of the CGS, while that of the the second is made to vary according to the processed flow rate. For this reason, the trend of the T_{water} of the CGS_B is extremely similar in terms of load profile to Q_{BU} and Q_{gas}. Considering system A, the temperature fluctuates for some periods, particularly around the turn of January, showing for water temperature the highest values, and then returns to standard values around 44 °C from May on wards.

Correlation analysis

In this section, a more in-depth analysis of the correlation between the variable that will be the target of our models, i.e., Q_{BU} preheat consumption, and all other quantities at play in the plant was carried out.

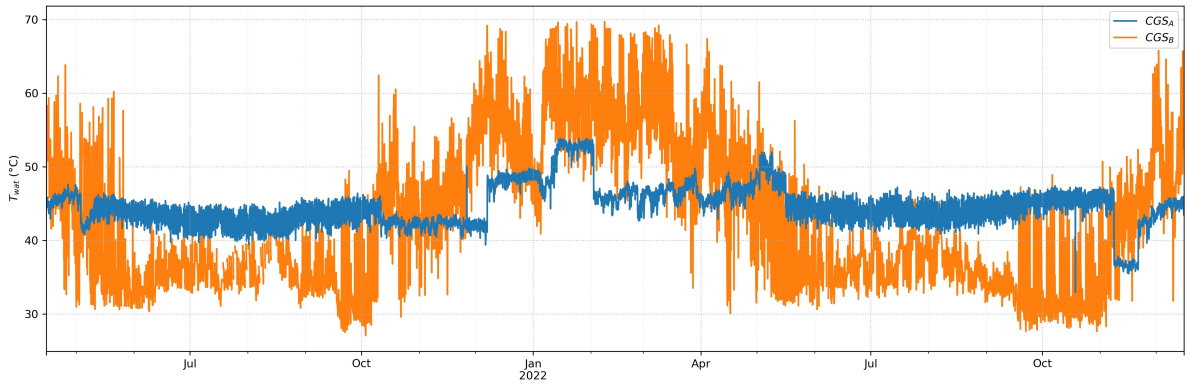
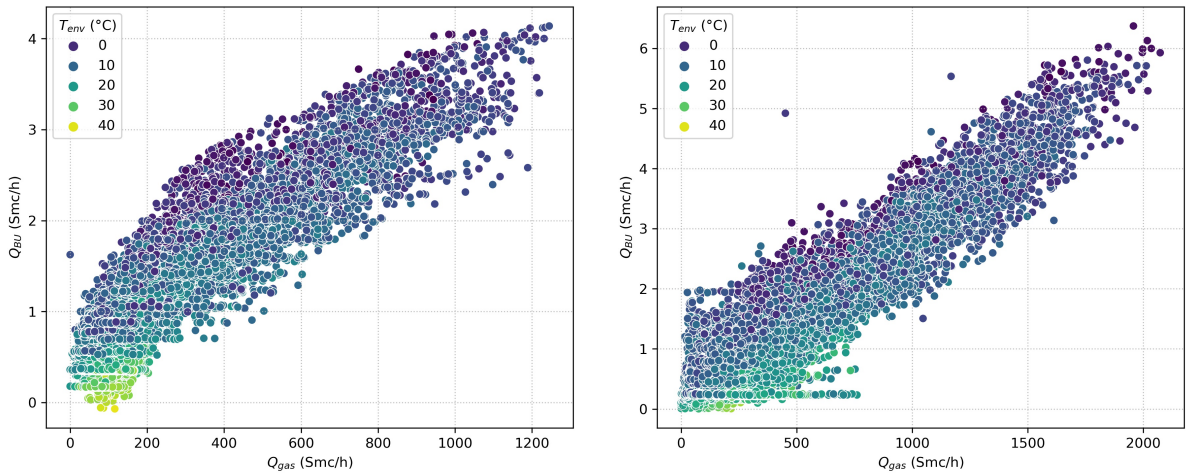


Figure 4.11: CGS_A vs CGS_B: preheating water temperature

To analyze these correlations we will make use of two different but complementary approaches. The first involves analyzing with a more physical and engineering approach the correlation between consumption and other parameters, and the second uses statistical-mathematical coefficients to evaluate and quantify the correlation between two variables. In this section, we will go into more detail on the first approach, which will be combined later with the second in Chapter 6, which deals with the implementation of machine learning models. Figure 4.12 shows the scatter plot of gas consumption values versus natural gas flow rate and environmental air temperature. In this graph we can see that there is a perfectly linear trend between the flow rate of gas consumed and the total flow rate flowing through plant B; moreover, the increase of Q_{BU} and Q_{gas} are clearly accompanied by the decrease in T_{env}. Different discussion for plant A: for a very low Qgas flow rate the trend is not linear but rather undergoes a kind of surge that stabilizes on a linear trend after a certain flow rate value, about 200 Smc/h. This is due to two combined factors: first, Plant A works with very low flow rates of natural gas and ReMi plants are always oversized in terms of thermal power plants, this could be significant that there is parasitic consumption due to the use of equipment in a low-efficiency zone because of strong load reduction. In addition, Plant A has 3 boilers instead of the canonical 2, as in the case of Plant B, and does not have a smart thermal power plant management system. These effects combined cause a nonlinear trend at the bottom of the graph, which results in an overall linear trend, but with a non-zero intercept and one that is around 0.5 Smc/h



(a) (b)

Figure 4.12: Scatter plot Q_{BU} vs Q_{gas} vs T_{env} for CGS_A (a) and CGS_B (b)

The third graph, ie the figure 4.14, is close to the first, only this time the variable z is the pre-

heating water temperature. It is notable a greater stratification of the variation in the pre-heating water temperature for plant A. While for plant B the T_{wat}. is modulated according to the Q_{gas} and therefore this correlation can be seen from the graph, for plant A if, for example, the vertical straight line is analyzed at about 400 Smc/h, it can be seen that regardless of the gas flow rate that passes through, consumption increases as the temperature of the pre-heating water increases, this means that in this plant some manual set point adjustments more manual to open from the gas flow rate Q_{gas} and which however effectively influenced the consumption of preheater Q_{BU}. The next plot of figure 4.13 is the union of the two previous graphs and was made to simultaneously compare the two systems.

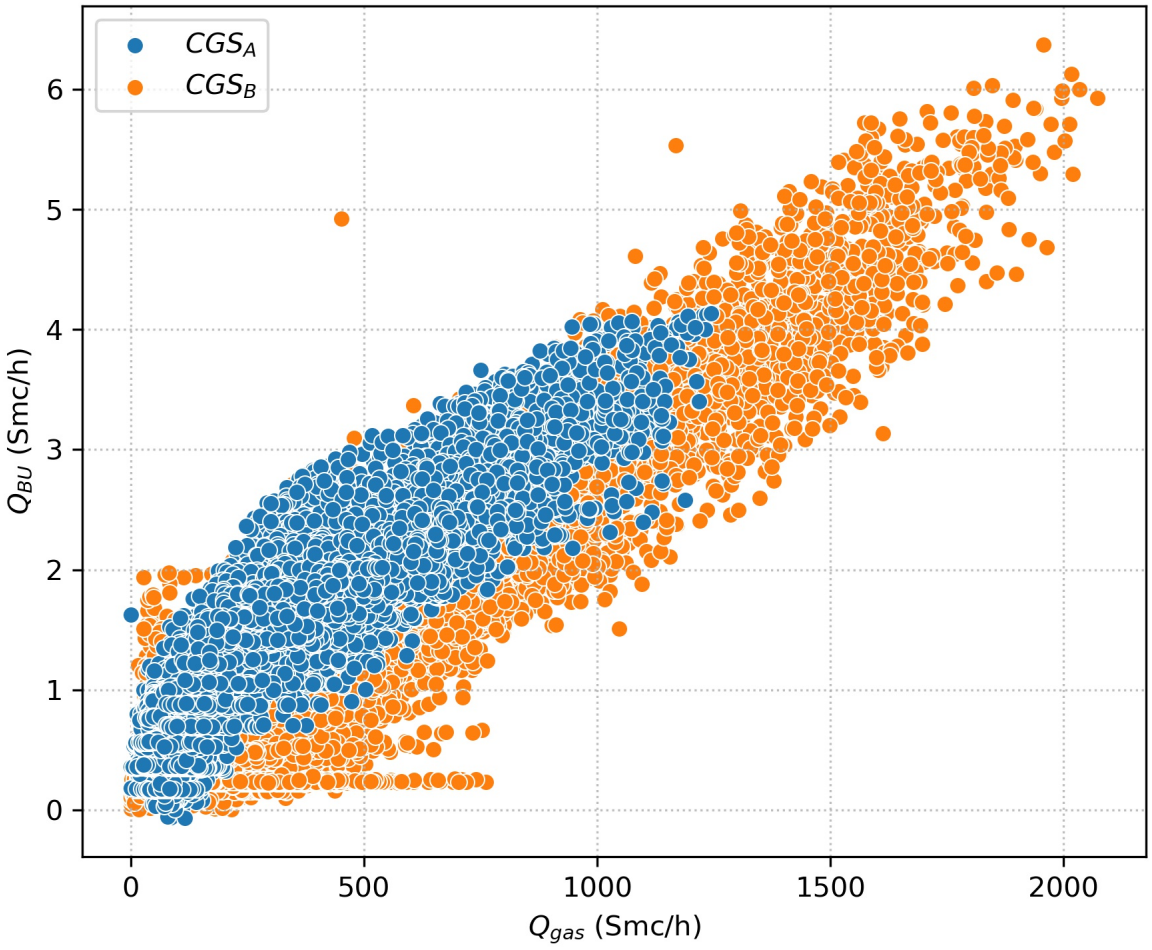
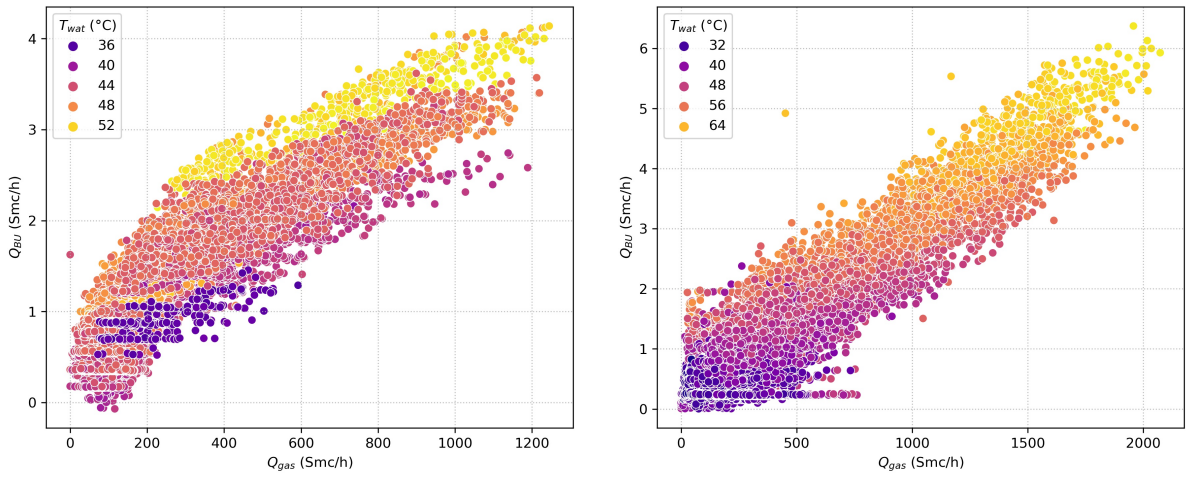


Figure 4.13: Comparative scatter plot Q_{BU} vs Q_{gas} vs T_{env} for CGS_A and CGS_B

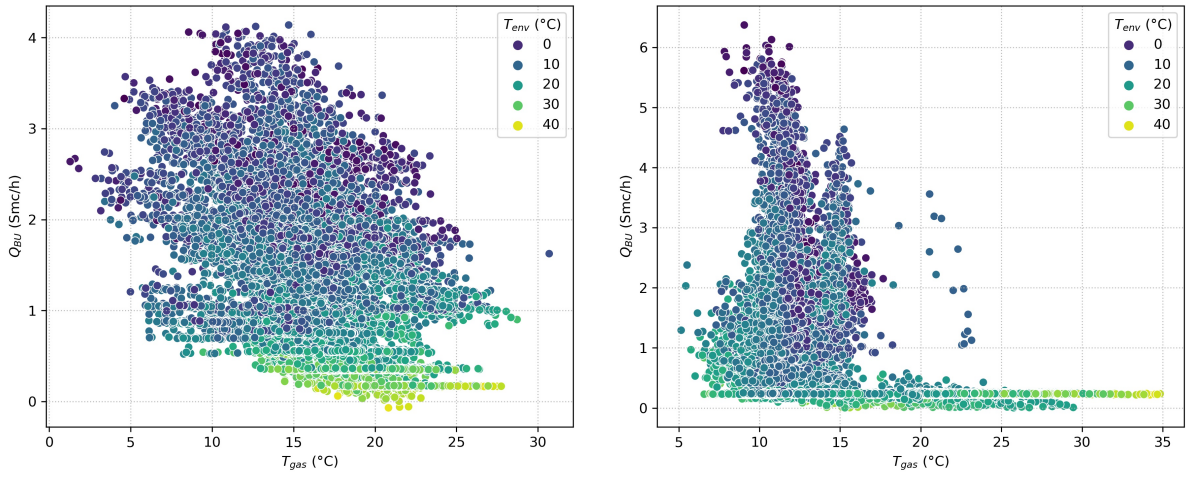
The fourth figure 4.15 shows the trend of Q_{BU} as the temperature of the gas leaving the CGS and the ambient temperature change. This time the scatters take on more scattered shapes and the correlation trend is more difficult to detect. For A, some inverse dependence of T_{gas} on temperature can be observed, thus a seasonal effect on gas temperature, while the dependence with Q_{BU} is weakly inversely correlated. Regarding plant B, two values of T_{gas} with a more pronounced frequency are noted, at about 12°C and 14°C, these could be two values related to the two set points of 8°c and 10°C but shifted forward a few Celsius degrees. Moreover, even for cabin B, there is a seasonal correlation, albeit weak, with the ambient temperature.

Finally, the last figure, i.e. 4.16, compares the consumption trend against the gas inlet pressure. The Q_{BU} is correlated with the inlet pressure, which in turn is physically correlated with the amount of gas that can be supplied to the CGS by the TSO. We see not only an increasing trend in consumption as



(a) (b)

Figure 4.14: Scatter plot Q_{BU} vs Q_{gas} vs T_{wat} for CGS_A (a) and CGS_B (b)



(a) (b)

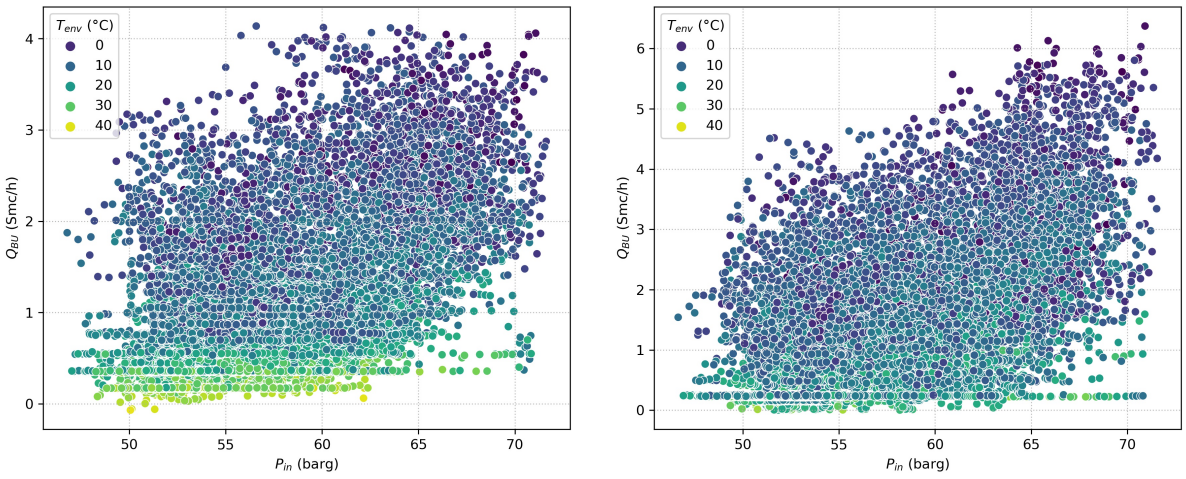
Figure 4.15: Scatter plot Q_{BU} vs T_{gas} vs T_{env} for CGS_A (a) and CGS_B (b)

P_{in} increases, but a more than linear trend. This is trend is identical in the modes for both plants. This is due to the dynamics of the gas transportation system from the TSO to the CGS makes it so that to pump more flow, a greater pressure jump must necessarily be overcome and thus there is this inherent relationship that nevertheless definitely goes to influence the correlation between Q_{BU} and P_{in}.

4.3.1 Thermal power plant control system analysis

Plant B, in addition to having a more advanced control system for the thermal power plant, was also set up to implement certain energy efficiency solutions foreseen by the DSO, in order to reduce the impact of the plant and gain access to Energy Efficiency Certificates (EECs) or "white" certificates [88]. During the calendar year 2022, the reactivation of a small heat pump was planned for the CGS_B cabin to partially handle the pre-heating demand of the plant. Figure 4.17 shows the course of the heat pump's timer, which assumes values between 0 and 60; this system subsequently had to be interrupted and was switched off by the DSO for repairs. Therefore, the system could only operate at full efficiency around the period of May-June 2022.

Figure 4.18 shows the variation of the outlet gas set point imposed by the DSO inside the CGS_B. This



(a) (b)

Figure 4.16: Scatter plot Q_{BU} vs P_{in} vs T_{env} for CGS_A (a) and CGS_B (b)

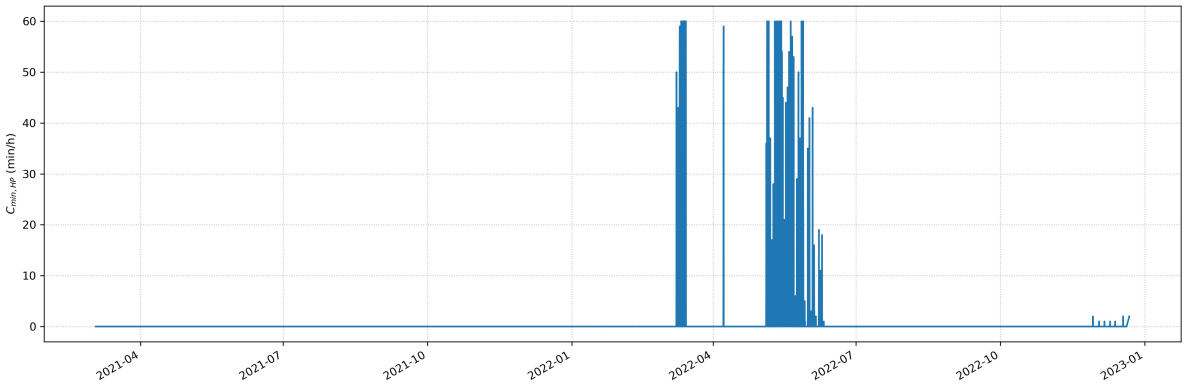


Figure 4.17: CGS_B: heat pump minutes counter

value is given to the feedback control system of the central plant of the CGS in order to matain the safety margin of the outlet gas temperature to avoid failures due gas cooling and hydrates formations problems for the machinery and pipes downstream the cabin. It can be seen that the set point may assume only two value: 8°C usually for the winter season and 10°C for the summer season. For the second summer under consideration (2022), the DSO decided to keep the gas set point at the CGS outlet lower and thus keep it at 8°C, instead of 10°C as planned and implemented for summer 2021.

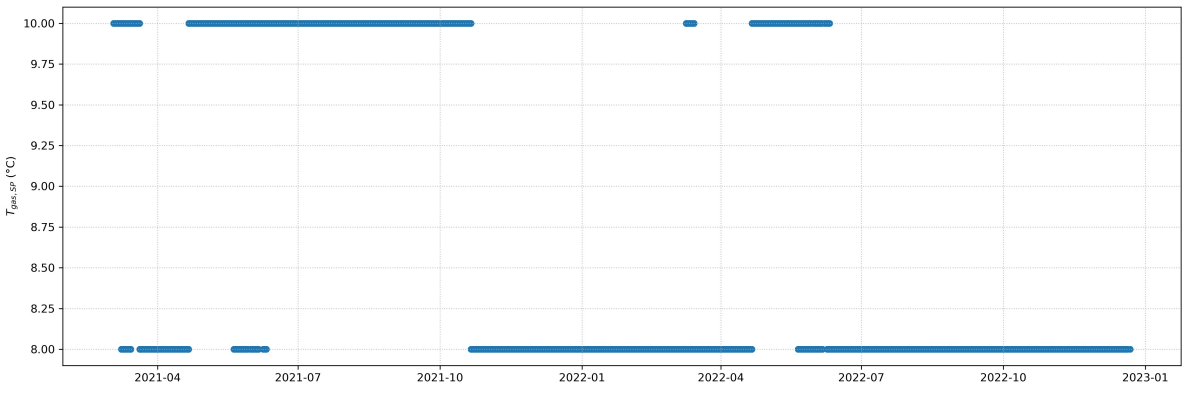


Figure 4.18: CGS_B: Outlet gas set point

Figure 4.19 highlights the trend of the water adaptive set point still for the second plant B. The system

is designed to follow the variation of the NG flow rates that pass through the plant and to adapt to that flow rate. It is so a value that is highly correlated with the natural gas flow rate and eventually with the natural gas preheating consumption flow rate.

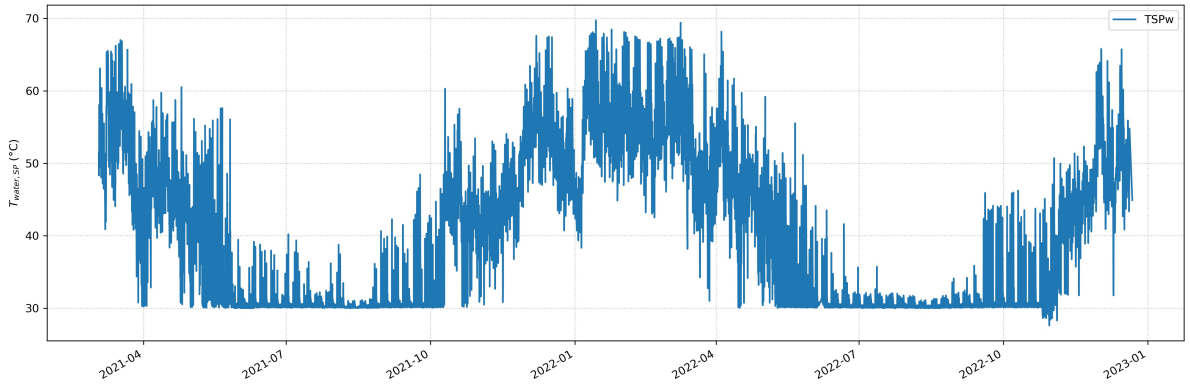


Figure 4.19: Preheating water adaptive set point for CGS_A

Figure 4.20 and Figure 4.21 show which Boiler unit is on or off and the number of minutes per hours of the overall system (60 minutes could be one boiler unit switched on for one hour or two boiler units switched on for half an hour at the same time or even in different moments of the same hour).

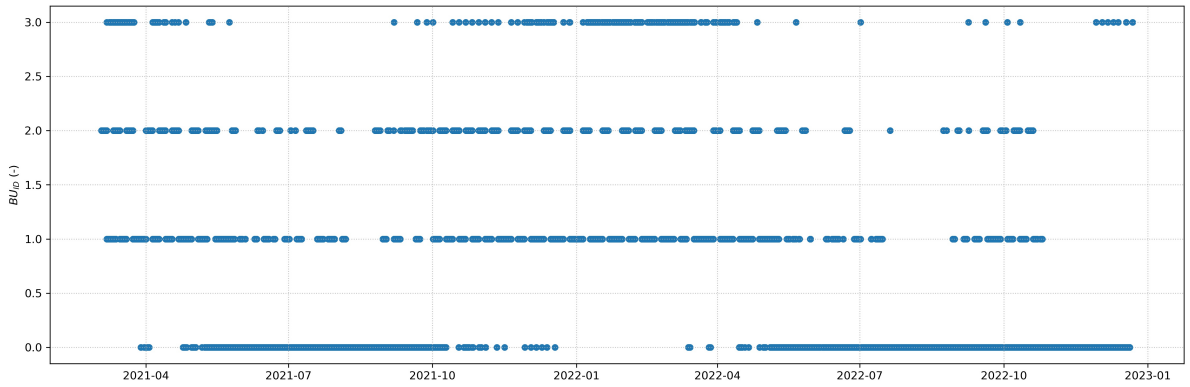


Figure 4.20: Boiler units ID: 0 for both switched off, 1 for BU_1, 2 for BU_2 and 3 for both switched on

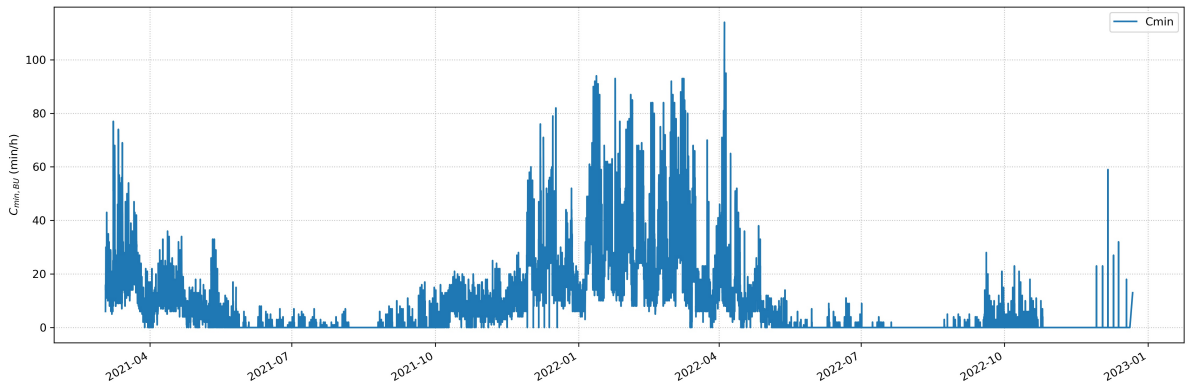


Figure 4.21: Boiler units minutes counter

In many cases, it was necessary to use the two boilers together, particularly during the winter period from November 2021 to March 2022. The peak of operation of the boilers can be seen from the minute's clock graph and particularly towards the beginning of April 2022, with a minute's clock value of more than

100 minutes/h, which means almost an hour of uninterrupted operation of two boilers at the same time.

4.4 Concluding remarks

These are the main takeaways from the dataset analysis:

- The daily NG consumption profiles of the CGS plants are slightly different while maintaining some shared peculiarities: consumption peaks are generally positioned in the same time slots. Knowledge of the type of users downstream of the CGS would allow us to extrapolate correlations between this and the type of profile accumulated at the CGS.
- There is a very strong correlation between NG consumed for preheating requirements and the main gas flow passing through the CGS. This was true for the only two data-set available, nut results suggest that this trend could be found for all CGS sizes.
- There is a strong seasonal variation regarding preheating consumption for each plant with marked differences between winter and summer thermal seasons.
- Systems complexity is difficult to generalize due to the disparate solutions that lead to different interactions between the quantities involved, such as outlet gas temperature, preheating water temperature, etc...

The cluster of analyzed CGS was selected by discussing with the DSO who provided the data set itself. The sizes of the different plants were chosen to have as representative a sample as possible of the framework of pressure reduction and measurement plants on the Italian territory. This allowed the same technological solution to be tested at various scales of required thermal power and processed natural gas flows in the next chapters. The group of CGSs is sufficiently representative in terms of the different flow rates processed and the pressure jump handled. Indeed, small- to medium-sized cabins are enormously more common than large and very large ones, which generally serve important industrial hubs or residential agglomerations in larger cities. It is also common to find multiple plants serving geographic points that require very high daily gas volumes and are less likely to have a single plant managing the total flow of a large city. The limitation of this cluster relates only to the lack of differentiation in the location of the systems and thus the downstream users belonging to climate classes. However, the plants considered are located in geographic areas related to the most densely populated climatic zones D and E and thus have the majority of gas demand in the Italian scenario. This cluster might be extended considering plants located in more extreme climate zones. This will affect the absolute value of gas demand at the same size as the reduction plant, but more importantly the shape of the gas demand curve. Greater annual consumption with different dependencies on seasonality may have an impact on the techno-economic feasibility of efficient plant solutions. Generalizing the treatment to different climate classes would be an additional step to add value to the analysis, but it may be hampered by the difficulty in sourcing data from various DSO companies, given the fragmentation of grid operators at the local level who control portions of the distribution network. In any case, this thesis's proposed method will retain its generality and may be extended to new plants.

Chapter 5

Results: techno-economic assessment for the decarbonisation of gas preheating

This chapter shows the results of the techno-economic feasibility analyzes for the decarbonization of CGS with RES-based HP. Preliminary results of preheating requests are shown first. From these, the PV and HP sizes on which to calculate the optimum are chosen and the analyses are carried out for three control logics, comparing three economic indices simultaneously: the NPV, the PBT, and the PI. This section presents the results of the analysis whose goal is to find the best techno-economic solution proposed by varying the size of the heat pumps and photovoltaic panels. The section is developed as follows:

1. First of all the calculations of the pre-heating requests are applied to each cabin and the cumulative curve of annual use is obtained which is used to define the size range of the HP to be analysed.
2. Then, plant n°1 is used to show the differences between the three proposed control logics in terms of energy production from the heat pump, energy sold and purchased concerning the network and so on.
3. Subsequently, the economic maps (NPV, PI and PBT) are shown for each solution and from these the techno-economic optimum for each plant is obtained. The optimum is then subjected to a sensitivity analysis to changes in gas and electricity prices.
4. Finally, the optimal solution found is analyzed from the point of view of the maximum percentage of decarbonisation obtainable and the conclusions are drawn.

Following this, the price sensitivity analysis is carried out and finally, the effective decarbonization of the plants is evaluated, considering the optimal techno-economic solutions found previously.

5.1 Technical analysis

5.1.1 Preheating requirements analysis

First of all, the reference period of the analysis is chosen: the 7 plant datasets are related to the period September 2021-September 2022, with hourly sampling. Figure 5.1 shows the daily value of the preheating energy requirements throughout the year selected for all the analyses while 5.2 shows the bar plot of the yearly preheating energy requirements for the 7 CGS analyzed.

The results of the analysis of the annual preheating request for each cabin are shown in the barplot figure 5.2: the span of preheating requirements in terms of MWh/year is from approximately 35 MWh for

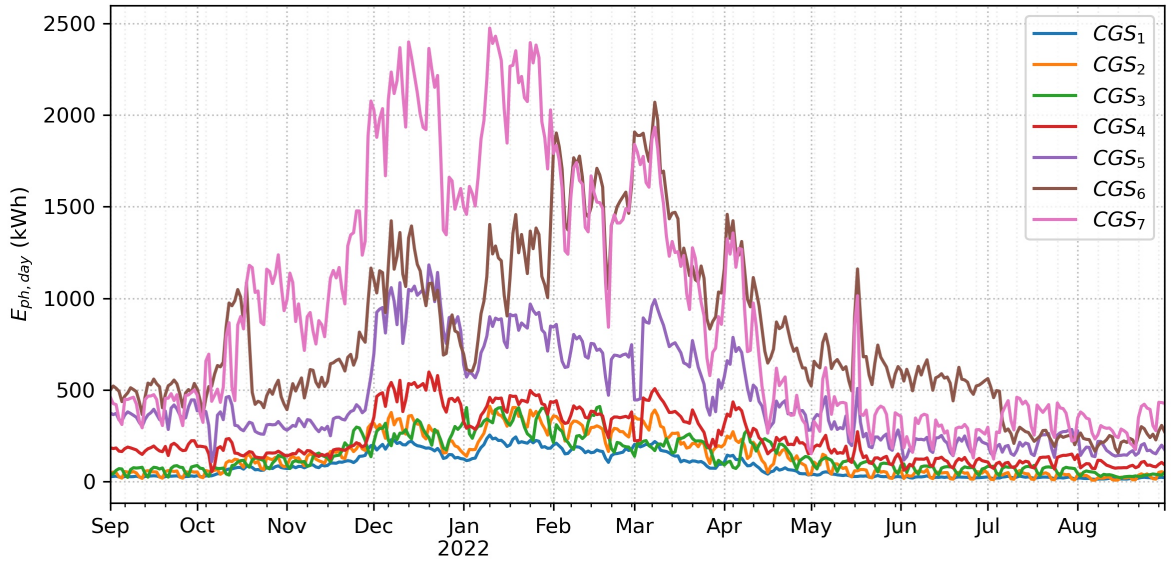


Figure 5.1: Natural gas preheating requirements for each plant every day.

the CGS_1 to about 350 MWh for the last plant, namely CGS_7. Within this cluster of plants, we have the first three that have assumed values of 35 MWh for the CGS_1, 50 MWh for both CGS_2 and CGS_3, with the latter showing a maximum value of Wph lower than the corresponding of former, but a higher demand plot on average throughout the year. From 4 to 6 there is a sudden growth in plant size (and therefore in $E_{ph,y}$ demand): about 90 MWh for CGS_4, about 175 MWh for CGS_5 and slightly less than 300 MWh for CGS_6.

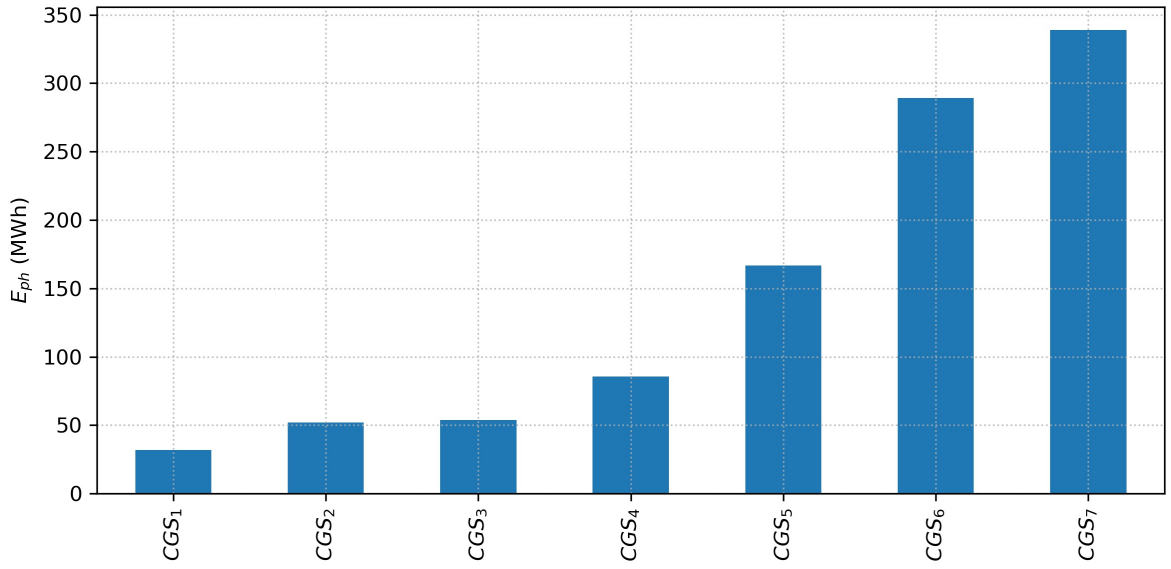


Figure 5.2: Natural gas preheating requirements for each plant yearly.

What is significant to highlight is not only the daily distribution (load curve of figure 5.1) or the annual total request for preheating, but also the values distribution of the hourly preheating powers for the CGSs 5.3. The flatter and more horizontal the curve is and the faster it drops from the maximum value to the average value, the more constant the pre-heating requests will be throughout the year; conversely, as occurs for example in the CGS_7, many hours are above a certain quite important pre-heating power value (30 kWth) considering the maximum value of 145 kWth. Comparing system CGS_7 with system CGS_6,

we see how the latter presents a much steeper descent from the frequency of the maximum power value required, immediately settling on medium-low values if purchased with the maximum. On the other hand, cabin 7 has a more homogeneous request variation between the maximum and minimum values. This will influence the techno-economic feasibility of the investment depending on the size of the HP-PV system.

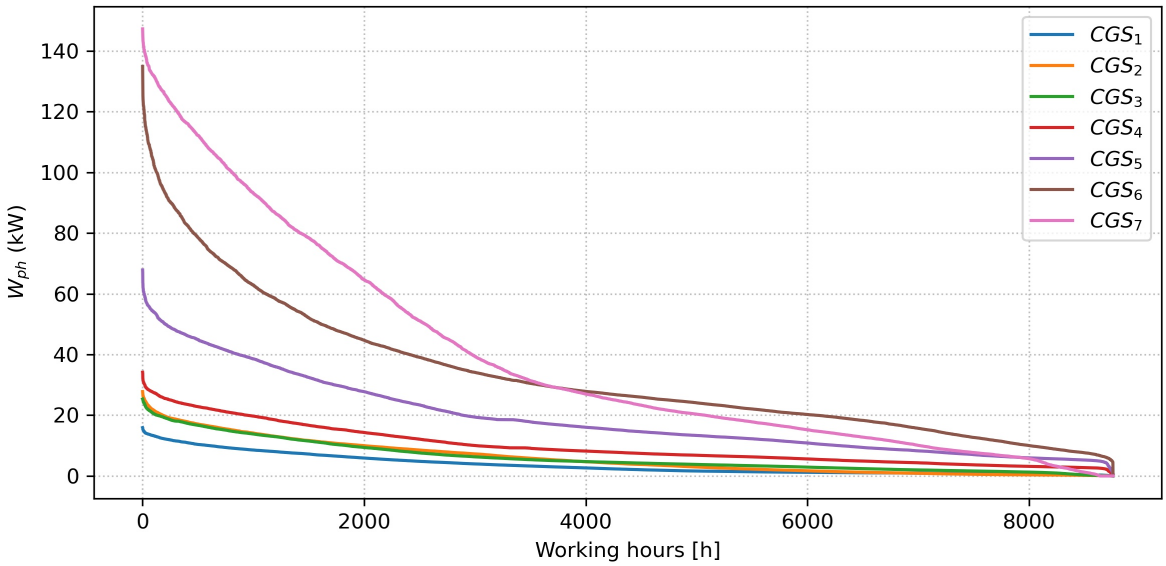


Figure 5.3: Natural gas preheating power distributions within the 8760 hours..

Knowing the distribution curve of the pre-heating power W_{ph}, the creation of the grid of pairs of sizes (HP, PV) was carried out to choose the sizing of the techno-economic optimum. The underlying hypothesis is to make a thermal power span of the HP from 25% of the W_{ph}, max for each plant up to 100%. The span relating to the size of the photovoltaic field was obtained by multiplying the span of thermal sizes of the HP by a seasonal COP, called SCOP, or (Seasonal Coefficient of Performance), which serves to give an average annual value of the COP of an ideal heat pump according to the place where they are installed. In italy a reasonable value for the SCOP is set to 3 [97], so the calculation for the i-th PV size is given by (5.1).

$$PV_{size,i} = HP_{size,i} \times SCOP \tag{5.1}$$

Figure 5.4 shows the percentage of gas that is consumed out of the total that passes through the cabin, it can be seen that it is around 0.1 % when approximated by excess for almost all plants. However, consider that the comparison is made between the ideal model and real gas flow rate, so the volume percentage may increase and even reach values of 0.2-0.3% depending on plant layouts, maintenance status of the plants, etc...

5.1.2 Control logic preliminary comparison

This section analyses the behavior of the hybrid system according to the different configurations of the control logic, described in the first section of the chapter. The cabin used for the analysis is CGS_1. The results of the first analysis are two-dimensional operating maps as the sizes of the heat pump and photovoltaic field vary; in the case of CGS_1, the span goes from 1 to 5 kWth for the HP and from 1 kWp to 15 kWp for the PV field. Figure 1.17 shows the heat maps related to the HP thermal production and the NG BU consumption for each control logic around the considered sizes of HP and PV. The map of the production of thermal energy in Figure 5.5(a) tells us first of all that as the size of the HP and the

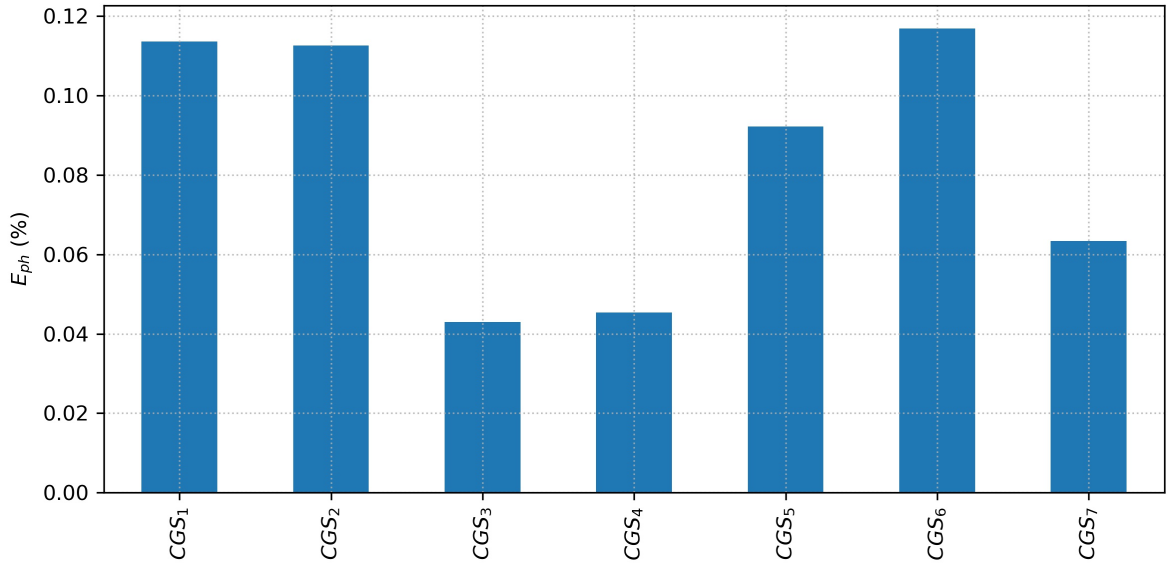
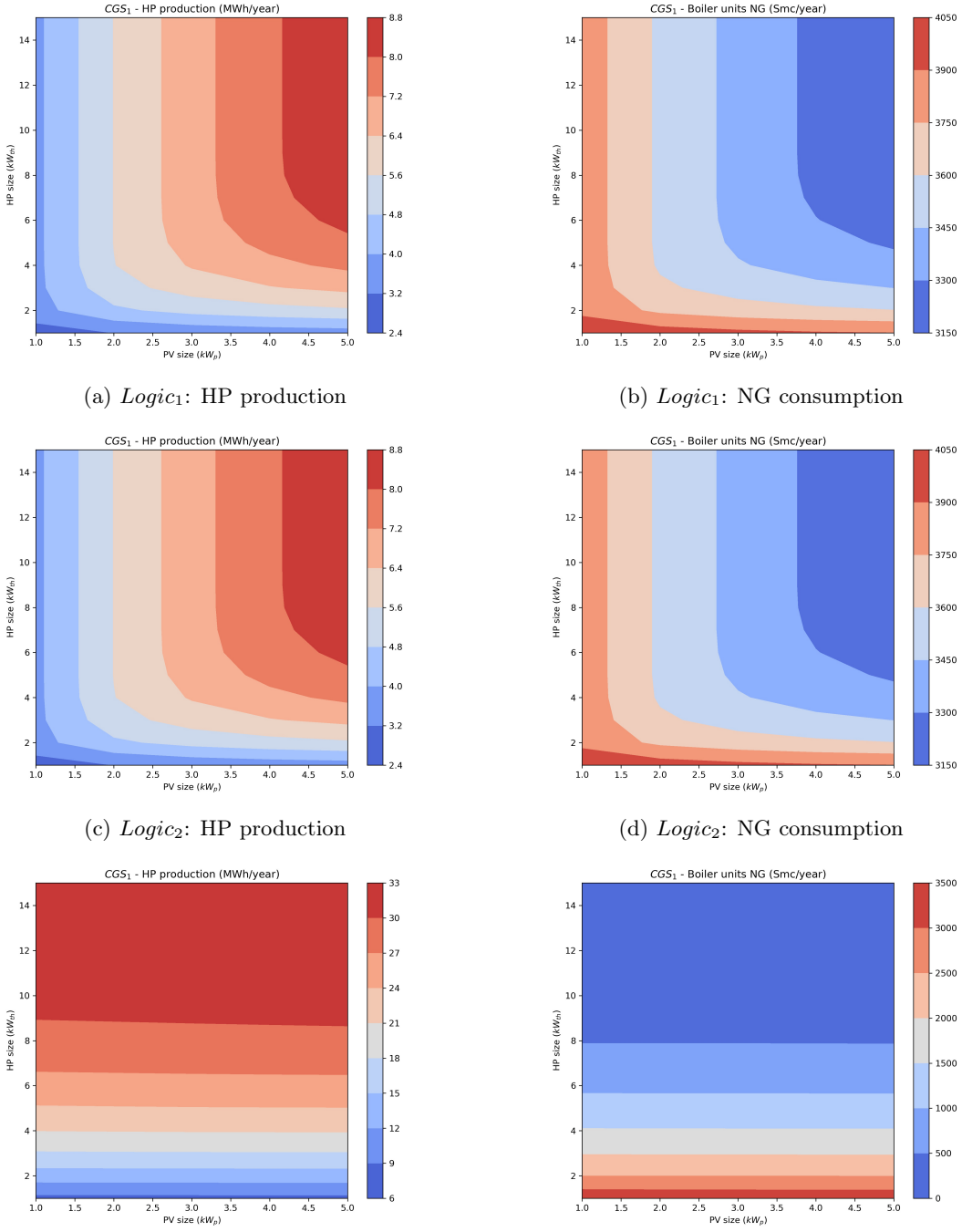


Figure 5.4: Natural gas preheating requirements respect to total yearly volume.

solar field increase, the power that can be produced increases. By analyzing Figure 5.5(a) it can be seen that with the same size of the PV, it is possible to produce up to a certain amount of thermal power with the HP, after which it is no longer possible since the HP does not receive sufficient electrical power from PV. The gas consumption map in Figure 5.5(b) is the mirrored version of Figure 5.5(a), as the size of the HP and the solar array increases, the natural gas that is burned decreases. Consequently, it follows that a combined linear increase in HP size and PV size does not lead to a linear increase in HP production, but there is a less than linear increase in HP production since as the PV size increases, one goes to work in areas that are always little accompanied by high preheat gas demands. This does not change for the second control logic which only avoids waste by selling the PV surplus on the network, while everything changes for the third control logic, in which, being able to buy and make up for the lack of PV, one goes exactly to match the production of thermal energy with the size of HP. For CGS_1, already with an 8 kWth heat pump almost 100% of the preheating request can be met, which in this case is about 33 MWhth per year; conversely, the use of natural gas for boilers goes to zero with the same trend. The second figure, i.e. 1.17, shows how moving from top to bottom (from off-grid logic to full on-grid logic) the maps are turned on such as selling on the network (c) and then buying from the network (e): in the case of off-grid logic both maps are off, in case of logic 2 (one-way on-grid) there is a sale of PV surplus energy to the grid according to the dedicated withdrawal price which grows with the PV size, while the purchase from the network decreases with the increase of the PV size and presents the maximum for the maximum HP size and minimum PV size. Finally, the relative maps of the SSR and the SCR, defined in the previous paragraph, are shown for the three control logics. The two parameters do not change as the percentage of self-produced energy (i.e. only from photovoltaic and not purchased from the grid) relating to the request W_{PH} is the same for the three logics, while for the SCR it is similarly evaluated how much of self-produced photovoltaic is consumed for PH. It can be seen from the SSR map that for cabin 1, which is used as an example of this work, a maximum value between 25 and 27.5 percent of the total energy required in a year for preheating was obtained. This number is indicative of the decoupling that is present between the natural gas demand curve and PV production. It means that in the absence of seasonal backup and storage systems, and the presence of a sufficiently powerful heat pump, it is only possible to reduce the carbon footprint of the system in off-grid logic by about 1/3. As for the SCR, on the other hand, the trend is mirrored against the SSR. The highest values of self-consumption ratio are obtained

for values of the lower Hp sizes regardless of the PV installed because in this case the SCR is defined concerning the thermal energy produced and does not take into account the high wastage of electricity confessed with a condition at small HP size and maximum PV size. The lowest SCR values are obtained for simultaneous high sizes of both HP and PV, standing at values around 30-40% for CGS_1 in this case.



(a) $Logic_1$: HP production

(b) $Logic_1$: NG consumption

(c) $Logic_2$: HP production

(d) $Logic_2$: NG consumption

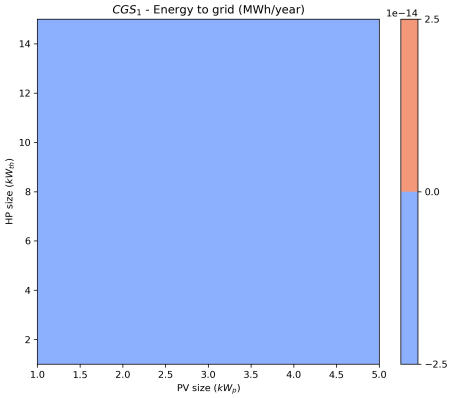
(e) $Logic_3$: HP production

(f) $Logic_3$: NG consumption

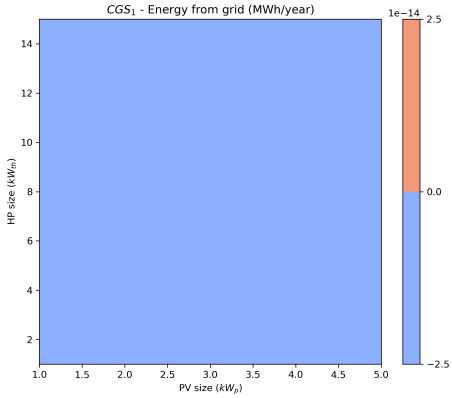
Figure 5.5: CGS_1: ASHP thermal energy production vs NG consumption from the Boiler Units for the 3 control logics.

5.1.3 SSR and SCR indexes assessment

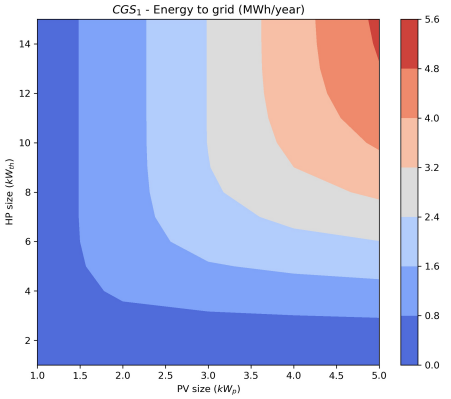
Figure 5.8 and Figure 5.9 show the results of SSR and SCR analysis for each cabin. The maximum values achievable by the system with off-grid control logic are: SSRs of 27.5 %, 36 %, 36 %, 30 %, 33 %, 33



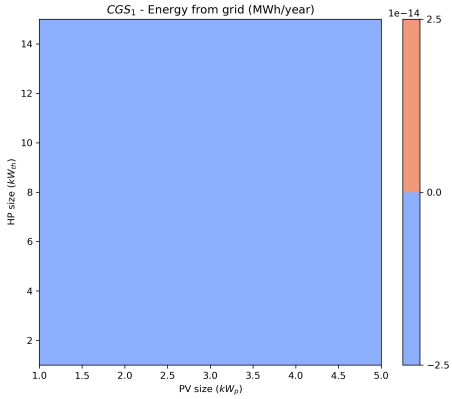
(a) $Logic_1$: energy to grid.



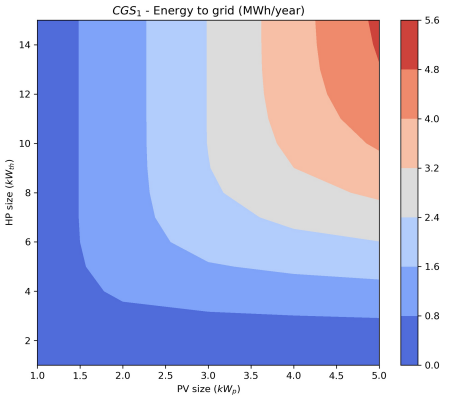
(b) $Logic_1$: energy from grid.



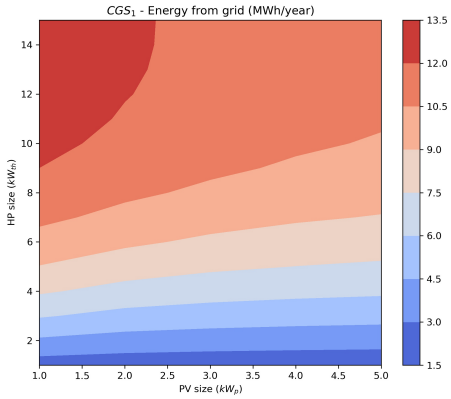
(c) $Logic_2$: energy to grid.



(d) $Logic_2$ energy from grid.



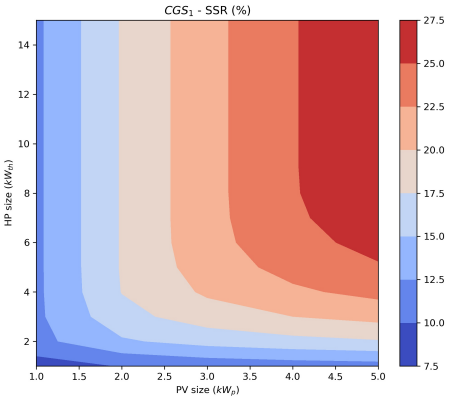
(e) $Logic_3$: energy to grid.



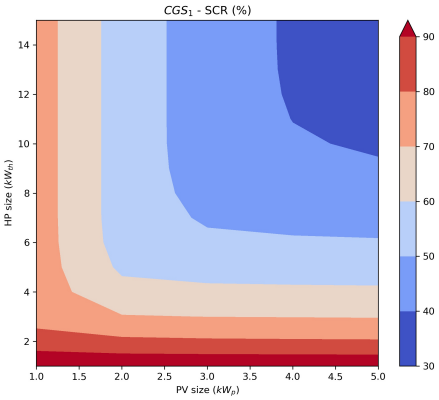
(f) $Logic_3$: energy from grid.

Figure 5.6: CGS_1: Energy sold to the grid vs purchased from the grid for the 3 control logics.

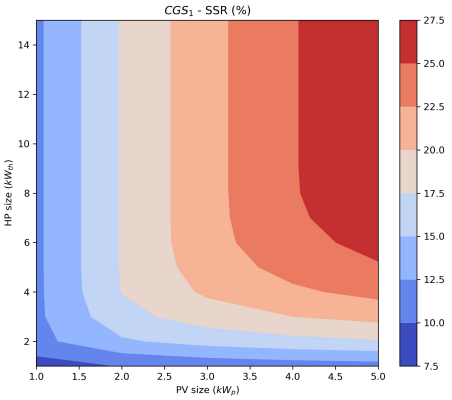
% and 32%. with SCRs of about 30 %, 32 %, 40 %, 40 % , 40 %, 40 % and 40%. The trend of the SSR index value follows fairly closely the trend of HP production in the previous figure, i.e., Figure 5.5. The SSR cannot increase beyond a certain value as the PV size increases because the system cannot use the surplus of electricity when there is no preheat demand. It is therefore extremely constrained to seasonal decoupling summer (high solar irradiance, little gas with preheat demand) and winter (less solar irradiance, lots of gas with preheat demand). CGS_3 and CGS_6 are the two cabins with the highest SCR on average. This can be seen from the values assumed by the colors in the respective figures at the top right, i.e., for high HP and PV sizes. The maximum achievable in these two cases ranges between 40



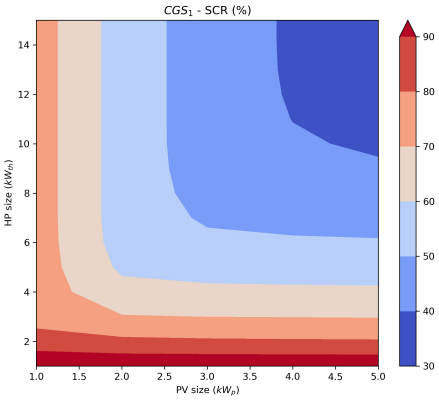
(a) $Logic_1$: SSR



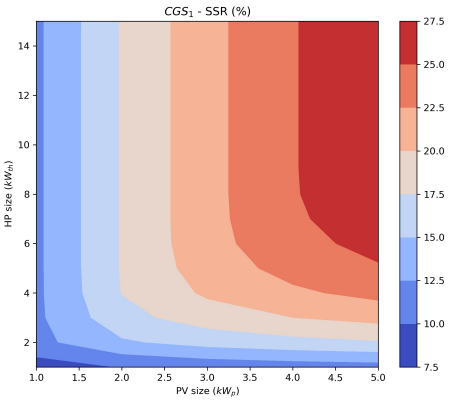
(b) $Logic_1$: SCR



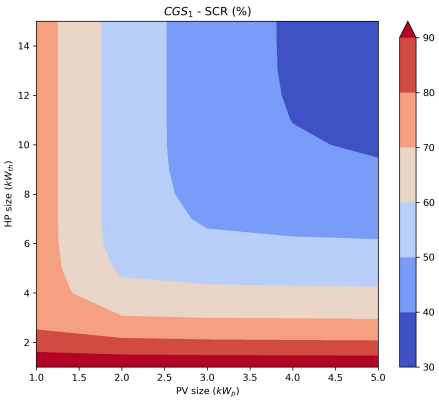
(c) $Logic_2$: SSR



(d) $Logic_2$: SCR

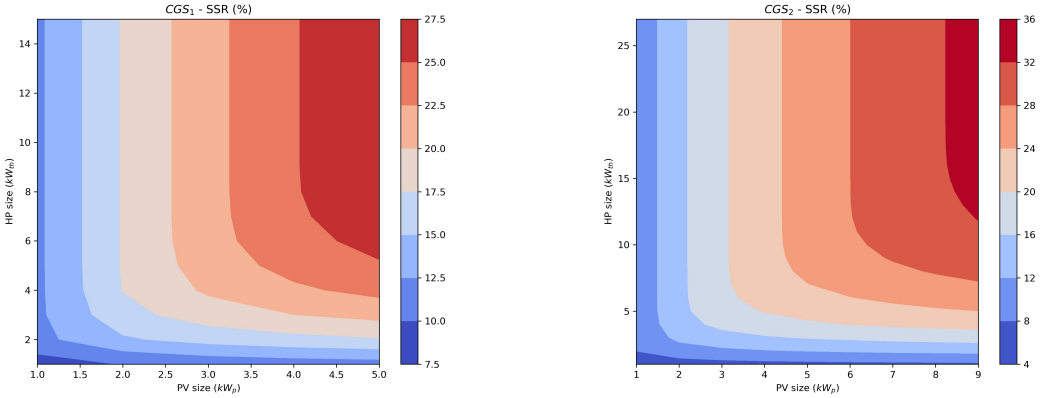


(e) $Logic_3$: SSR



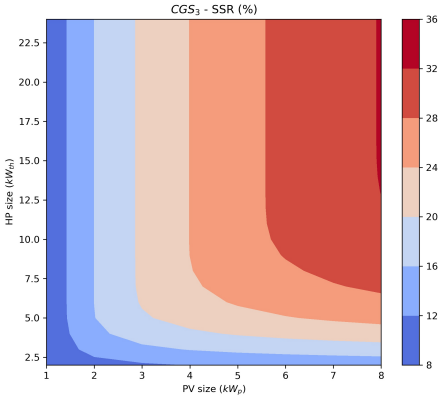
(f) $Logic_3$: SCR

Figure 5.7: CGS_1: SSR vs SCR for the 3 control logics.

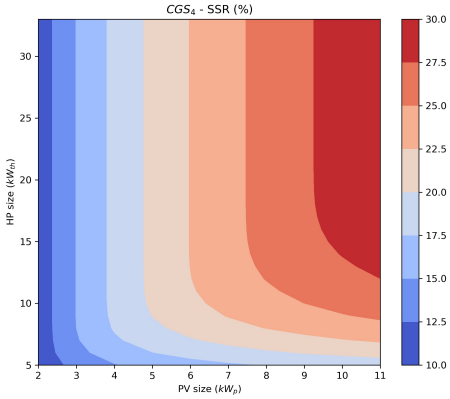


(a)

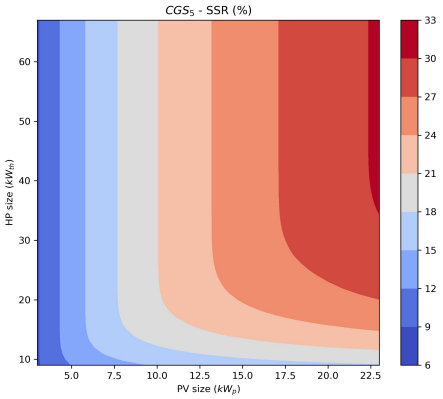
(b)



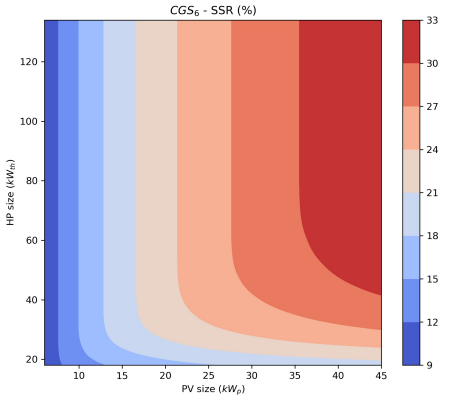
(c)



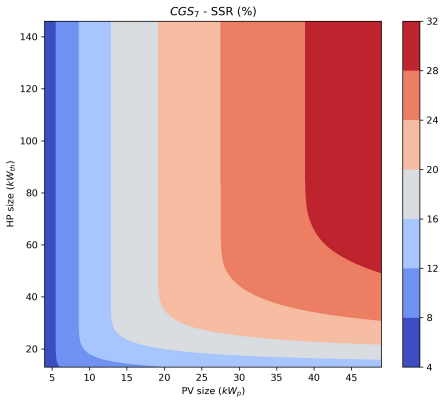
(d)



(e)



(f)



(g)

Figure 5.8: SSR results for all plants with control logic 1

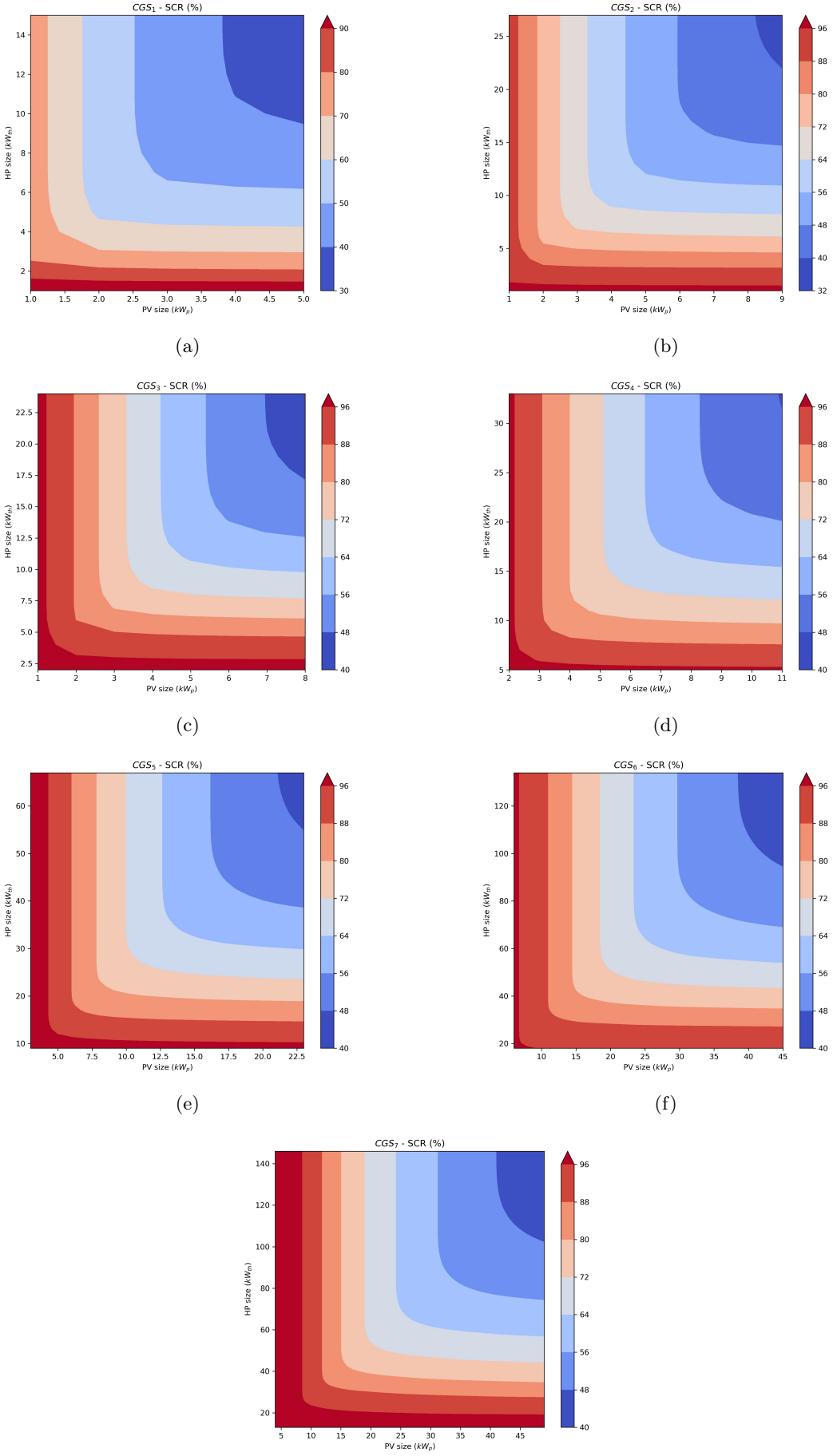


Figure 5.9: SCR results for all plants with control logic 1

5.2 Techno-economic assessment

This section analyzes the economic index maps (PBT, NPT and PI) for all plants, number to number 7, and for each proposed control logic from off-grid to fully on-grid. In this first part of the paragraph we will analyze the maps of the NPV (figure 5.10), PBT (figure 5.11 and PI(figure 5.12) obtained using logic 1, ie the off-grid logic. Analysis of the NPV maps show us that the investment configurations leading to a higher 20-year NPV are found around medium-high PV field sizes and medium-low heat pump sizes: this is because that with medium to large PV fields, there is maximum electrical production and sale to the grid, and thus the possibility of exploiting larger heat pumps. On the other hand, however, if the payback time and profit ratio are also considered, the techno-economically optimal solutions for the first control logic move into the area of small to medium sizes for the two technologies simultaneously. Thus, considering the three indices as a whole shows that for a totally off-grid configuration:

- incresing the size of PV pays more only up to a certain size of heat pump, after that size the investment is no longer profitable. This is due the fact that the maximum amount of saved NG is reached and after that particular point, the gain due to saving more gas because of the increase in PV size and HP slows down a lot because of the decoupling of the demand curve with the PV production curve.

- The area of the map (HP,PV) that never allows it to be profitable is as to be expected the one with very high heat pump size and very low PV size; this is evidently because that the pump is extremely oversized compared to its electric counterpart.

- If for PBT and PI the trend of the maps is almost identical, for the NPV map it is noticeable how the CGS_A presents a different trend instead: the zone of highest profit at 20 years is no longer at high PV sizes but at medium-low sizes. It means that plant 1, which is affected more by fluctuations due to its size, in fact it is extremely small in terms of volumes of gas processed annually, presents a point of economic optimum for sizes smaller than the pair (HP,PV). Therefore, a small-scale intervention allows the point of approximately higher savings of natural gas at lower cost to be sustained. Beyond that point the intervention already begins to become less profitable.

Now we will analyze the maps of the NPV (figure 5.13), PBT (figure 5.14 and PI(figure 5.15) obtained using logic 2, ie the on-grid logic with only the selling direction to the grid. In this case, you allow the PV system to sell the surplus energy back to the grid. This was considered because , while we do not want to transform these systems from gas-fired reducers of consumption to electricity producers, it certainly helps the investment to reduce PBT and increase profits by selling the surplus back to the grid. The NPV maps, if compared with those of the previous control logic, do not show a substantial increase in terms of maximum absolute values of NPV, but rather an enlargement of the range of sizes which allow to obtain the highest NPV values after twenty years. Furthermore, the slope of the front with which profitability is reduced rises for all plants: it means that the increase in PV allows a more positive economic return compared to the off-grid case alone. Even the PBT and PI maps do not undergo substantial variations, the range of greater profitability also widens for these and consequently the range that does not lead to a return on investment during the useful life of the plant is reduced. For the third group of plots we will analyze the maps of the NPV (figure 5.16), PBT (figure 5.17 and PI(figure 5.18) obtained using logic 3, ie the on-grid logic with the selling direction to the grid and also the possibility of electricity purchase. For the NPV it can be seen that the transition to the totally on-grid control logic almost eliminates the cases in which the NPV is negative. Using the network as a backup allows you to use the HP for thermal purposes from 40 to 60% more than in the totally off-grid case. Furthermore, the areas of

excellent economic performance from the point of view of the NPV are for medium-high HP sizes, but not maximum, and for high PV sizes. This is because that, regardless of the annual cabin demand curve, a medium-high HP power often allows even more than 3/4 of the preheating request to be covered. If one remembers the violin graphs of chapter 2 and the preheat graphs of the beginning of this chapter, it will be deduced that the last HP sizes (the very high ones) are related to very rare occurrences. Due to the statistical distribution, the NPV basically stands at around 75% of the preheating power for the various substations. The payback times are lowered by an average of 2-3 years compared to the two previous cases and the optimal areas widen considerably, while the areas where the investment is almost never convenient are reduced. For PBT and IP, however, the areas with the highest values for medium-low PV sizes and medium or medium-low HP sizes remain.

To summarize what can be deduced from the analysis of the graphs just described and reported:

- The optimum NPV tends to shift from the bottom left corner (low HP and PV sizes and therefore low Capex investments and small Opex that recover little energy) to the top right corner (high HP sizes and PV, such as important investments that aim to recover almost all the gas regardless of whether the energy source that feeds the HP is 100% renewable or only partially renewable).
- The PBT value has reached a minimum of 6-8 years for all plants for off-grid logic and one-way on-grid logic, while it is possible to recover the investment even with a minimum of 3-5 years if one chooses logic n° 3.
- The profit index PI on the other hand, keeping its optimum in the same position for all three control logics, i.e. the maximum that can be extracted proportionally to what has been invested, is obtained for medium-small size plants, i.e. spending little but being sure of recovering a lot of natural gas in proportion to the amount invested.

The pairs of optimal sizes have been considered for all three indices, but giving precedence to the actualized payback time, as this is favorable for the DSO who needs an economic payback time as short as possible, but is nonetheless interested in accessing incentives.The following table 5.1 summarizes the pairs of optimum sizes for each substation and for each control logic. These sizes of techno-economic optimum will be subsequently tested through a sensitivity analysis to gas and electricity prices and the results are presented and analyzed in the following paragraph.

Table 5.1: Summary table of optimum pairs (HP, PV) for each plant and logic.

Logic	CGS_1	CGS_2	CGS_3	CGS_4	CGS_5	CGS_6	CGS_7
Logic 1	(2,1)	(5,3)	(8,5)	(8,5)	(19,11)	(30,20)	(34,22)
Logic 2	(2,1)	(7,5)	(8,5)	(8,6)	(19,11)	(32,24)	(34,22)
Logic 3	(2,1)	(2,1)	(6,3)	(10,3)	(20,9)	(40,12)	(34,12)

5.2.1 Sensitivity analysis

This section analyses the results of the sensitivity analysis of the techno-economic optimum for each plant and for each analyzed control logic to changes in gas prices and the relationship between the gas price and the purchase price of electricity. In this analysis, however, the resale price recommended by the system operator is kept constant. The sensitivity analysis was carried out as follows: for each plant and each control logic, different natural gas prices are analyzed simultaneously (from 1.6 €/Smc to 0.4 €/Smc and

for different price ratios between gas and electricity equal to 3,4 and 5). This allows to understand how sensitive the investments made with techno-economic studies are.

Control logic 1 and 2

For the two control logics 1 and 2, the sensitivity analysis is with only one degree of freedom, i.e. only when the gas price varies, since the purchase price for the first two control logics is not influential on the final techno-economic optimum result. The sensitivity analysis for the first two control logics showed that the techno-economic feasibility of these solutions is very susceptible to gas prices. The first two plant logics (no purchase from the network) are extremely penalized if the price of gas settles on pre-crisis values and is lower than around 0.6-0.5 €/Smc. Therefore, halving the selling price of gas, other conditions being equal, such as the gas/electricity price ratio, is the maximum limit beyond which the investment does not become profitable even after 20 years. or its useful life. As regards the payback time, it is still possible to recover the investment in about 10 years if the gas price is equal or higher than 0.8 €/Smc. Results are shown in figure 5.19.

Control logic 3

Regarding logic 3, it is necessary to examine three graphs per cabin, as this time the sensitivity analysis is carried out on two degrees of freedom because now the price ratio affects the final result. Figure from 5.20 to 5.26 show the trend of the sensitivity analysis as the case study varies from CGS_1 to CGS_7. For all plants, it can be seen that a reduction in the price ratio between gas and electricity is detrimental to the resilience of the investment to price variability in this sense. One plant out of 7, or the CGS_4, as can be seen from the figure 5.23(a) never manages to recover the investment for any value of the price of natural gas. From the various figures it can be seen that plants 4, 5 and 6 are the most susceptible to a reduction in the price ratio, this is because proportionally they are the three plants whose configuration (HP, PV) relies more on the electricity purchased from the network in the case of techno-economic analysis. A scenario that envisages instead a greater increase in the price of gas than that of electricity and hopefully because of the increasingly greater decoupling between the energy produced from gas and the energy produced from RES, means that the investment acquires profitability. In that case, it is possible to reduce the PBT of the investment by about 1-2 years depending on the substation at the same price of natural gas or it allows to resist even a significant change in the price of gas. In the worst case (gas at 0.4 €/Sm3), a PBT of less than 7 years would be obtained for some substations and even less than 5 for others.

In conclusion, the off-grid solution and the partially on-grid solution have less impact from an economic point of view and allow the use of only renewable energy, but present a marked variability with the price of gas. At the same time, the solution that provides for the possibility of purchasing from the network guarantees a much greater resilience of the solution and its technical-economic feasibility as gas prices vary with a control logic 3. This is mainly due to two factors: reduction of the gas price is somehow balanced by the fact that a significant share of gas is still saved by purchasing electricity at a lower cost and producing an amount of heat on average three times the amount of electricity purchased and because, the ratio would also increase the gas and electricity prices, favoring this solution even more. The only scenario in which even logic 3 would not allow a sure return on investment would be a scenario in which the prices of gas and electricity are more coupled, which is decidedly unlikely given the prospects of ever-greater diffusion of renewables and the increasingly insistent debates on the decoupling of these two methods of electricity production (see thermoelectric from natural gas and photovoltaic) on the two correlated prices.

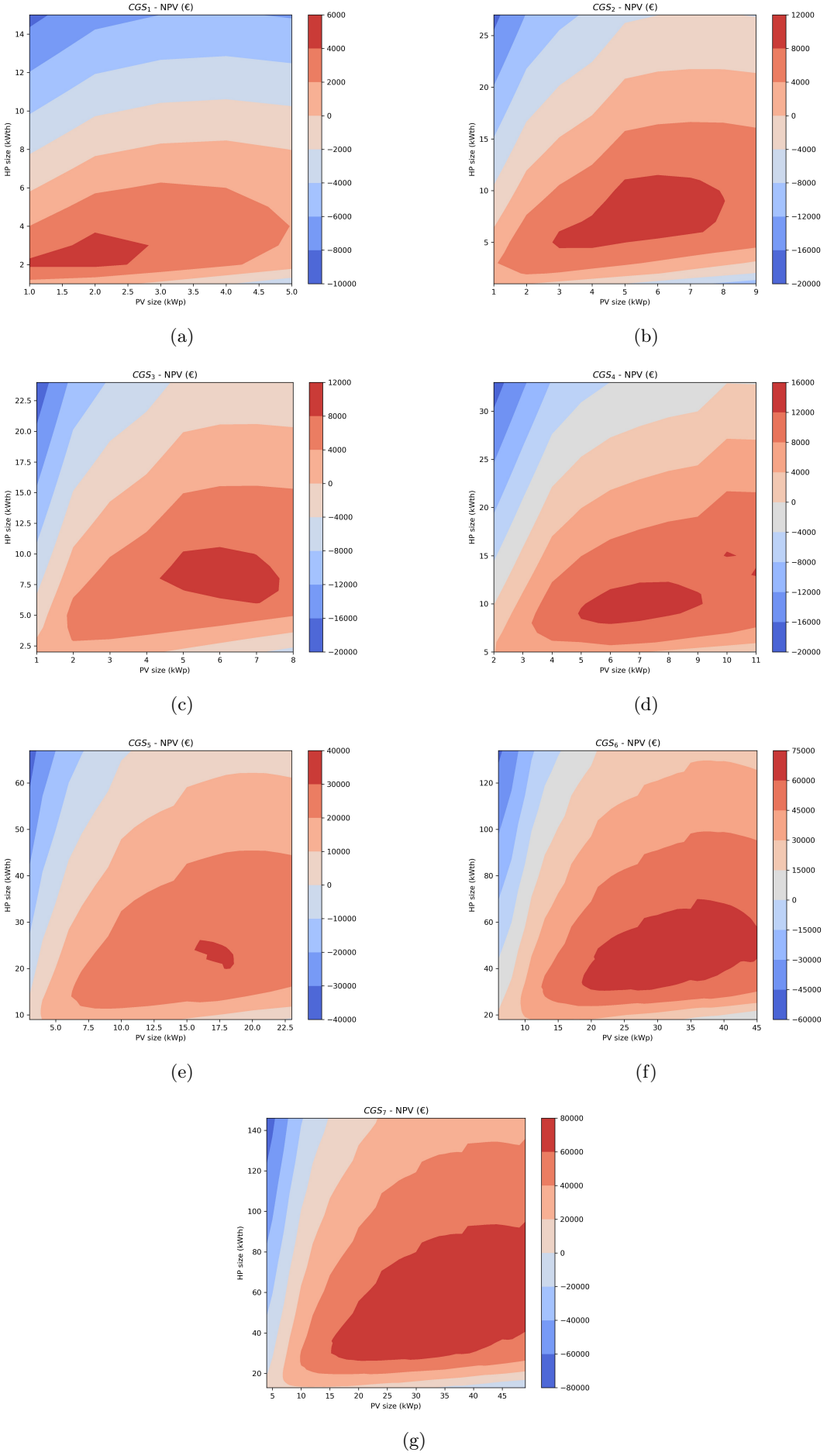
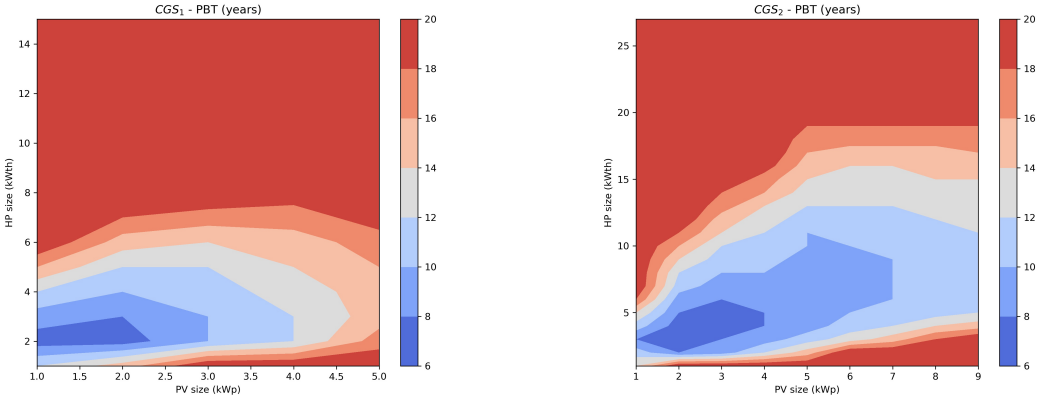
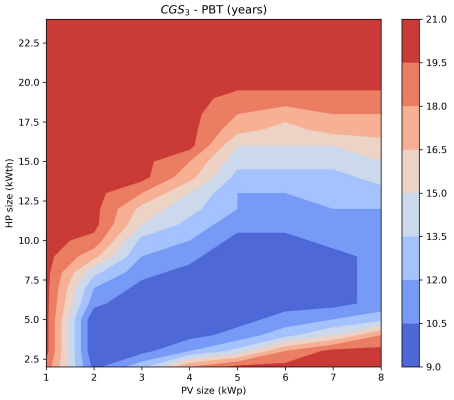


Figure 5.10: Net Present Value with control logic n° 1

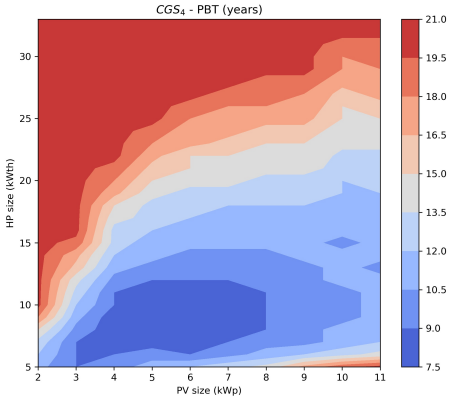


(a)

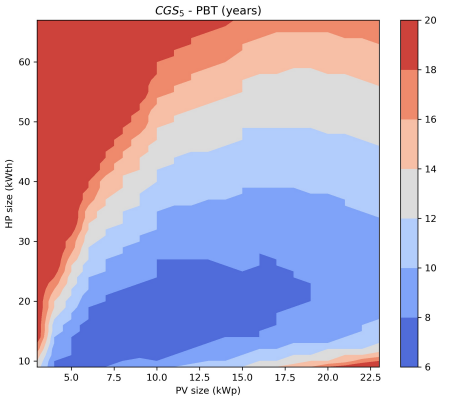
(b)



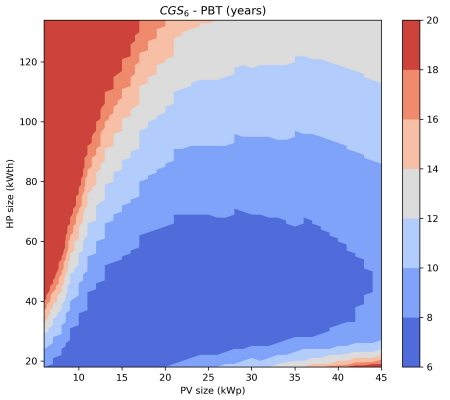
(c)



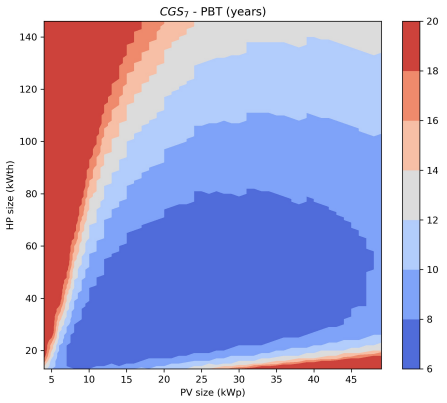
(d)



(e)



(f)



(g)

Figure 5.11: Pay Back Time with control logic n° 1

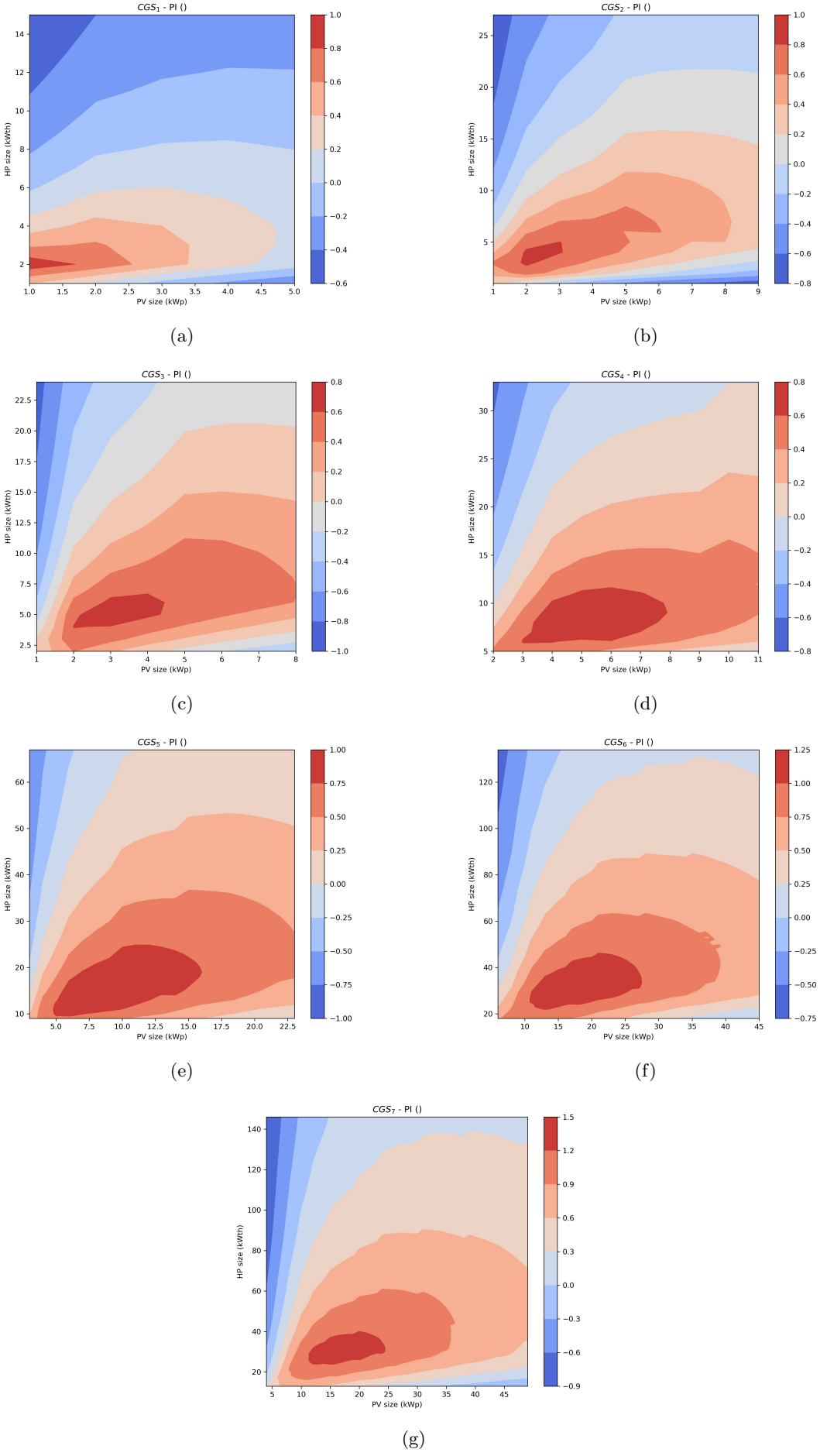
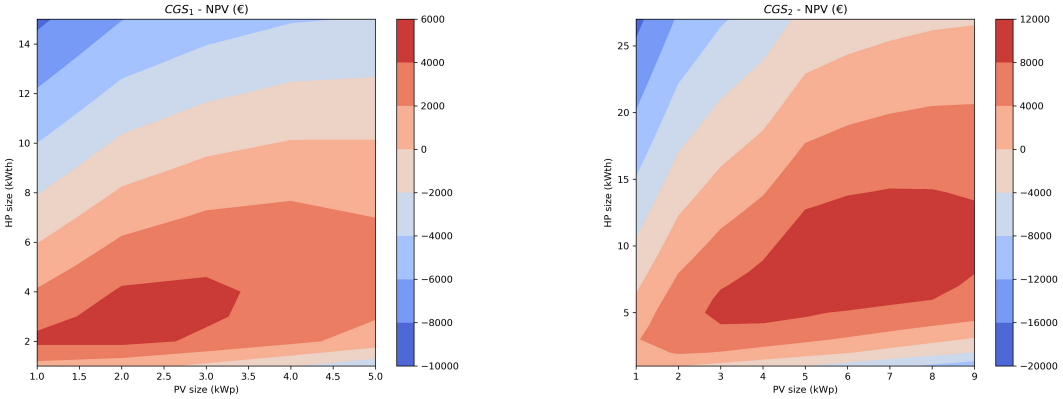
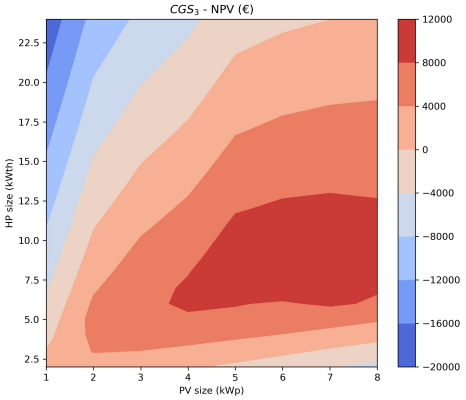


Figure 5.12: Profit indexes with control logic n° 1

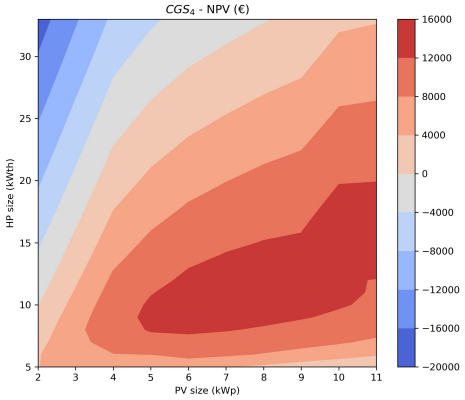


(a)

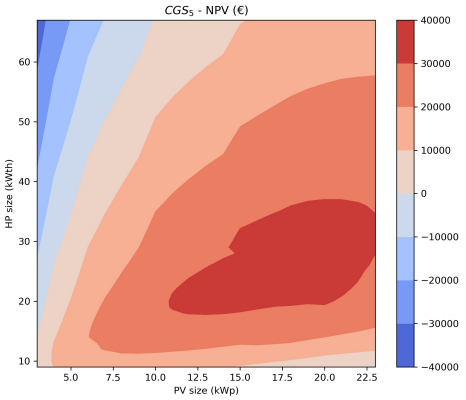
(b)



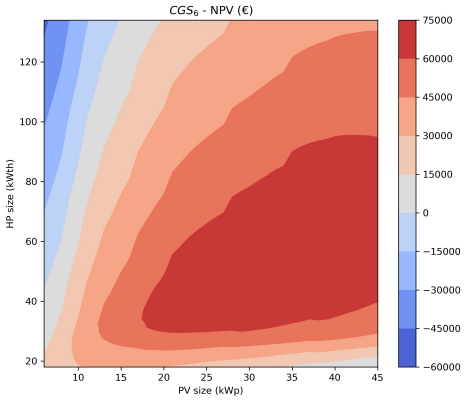
(c)



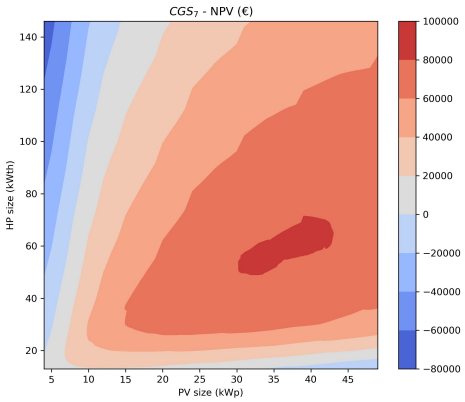
(d)



(e)



(f)



(g)

Figure 5.13: Net Present Value with control logic n° 2

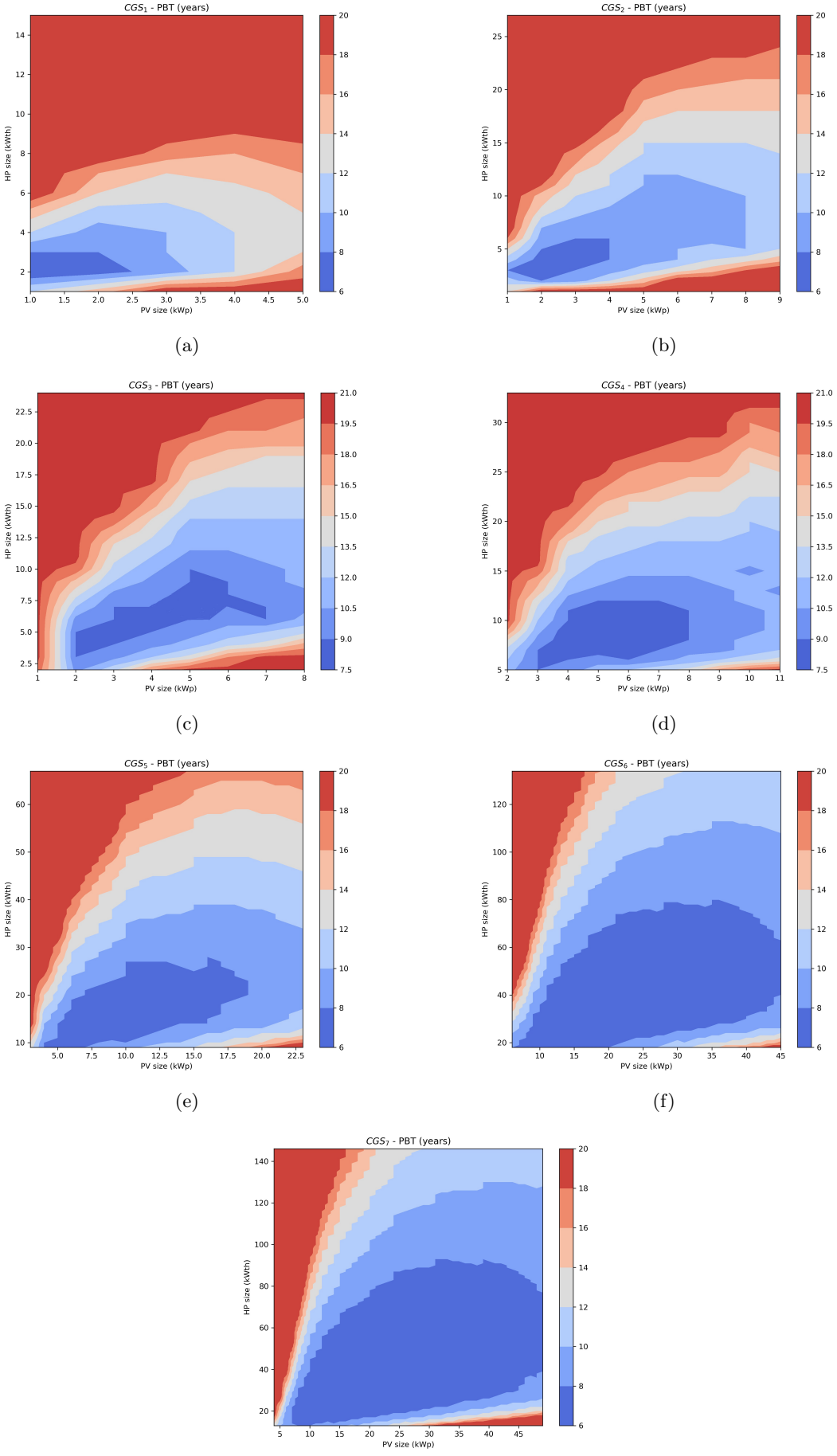
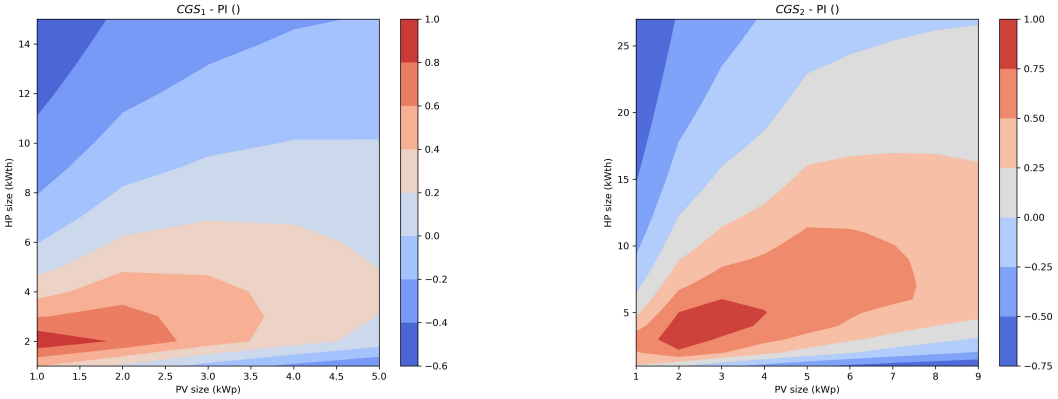
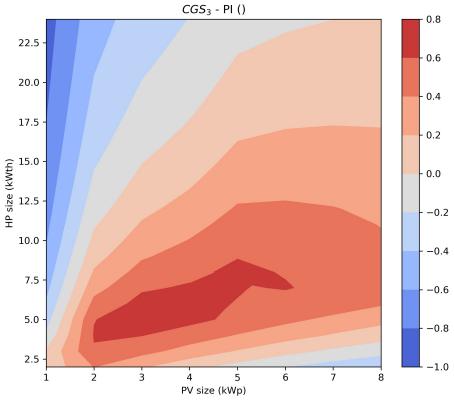


Figure 5.14: Pay Back Time with control logic n° 2

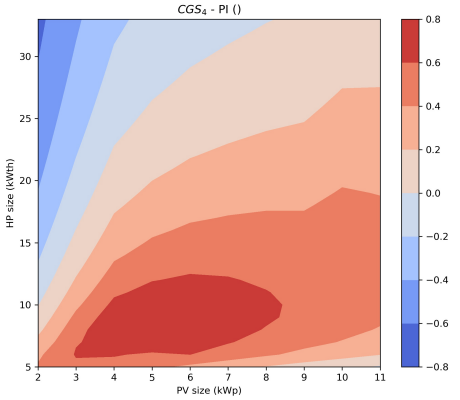


(a)

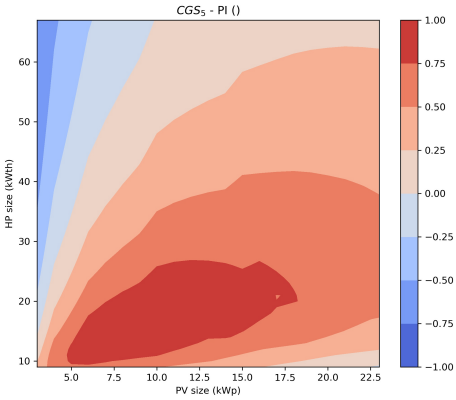
(b)



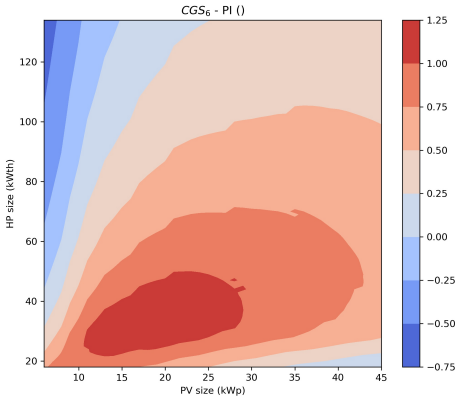
(c)



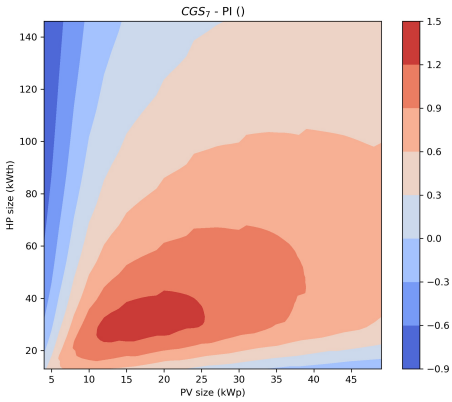
(d)



(e)

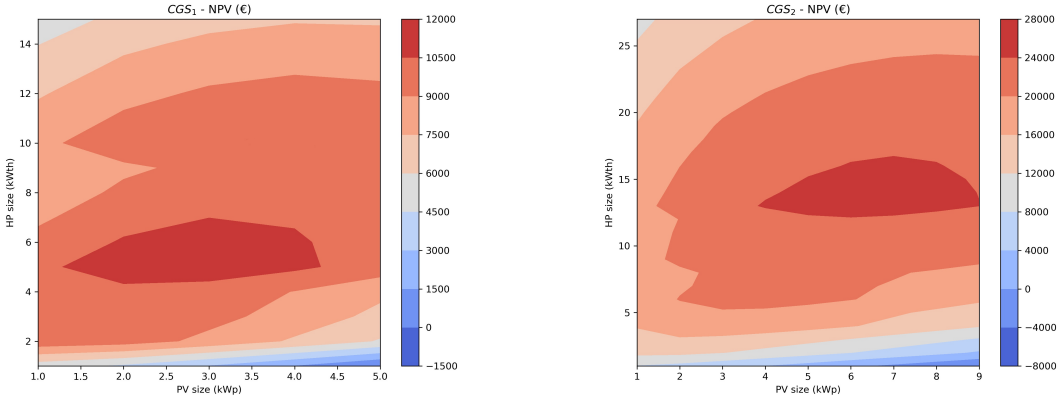


(f)



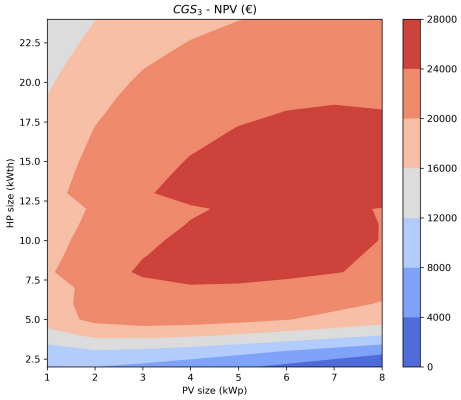
(g)

Figure 5.15: Profit indexes with control logic n° 2

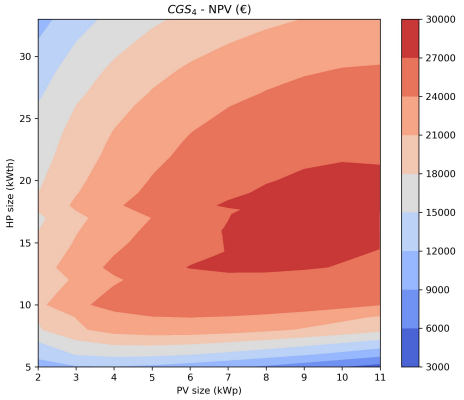


(a)

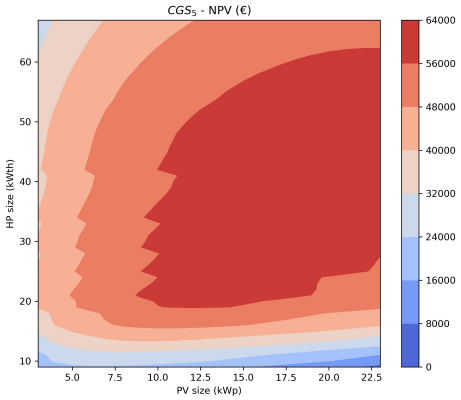
(b)



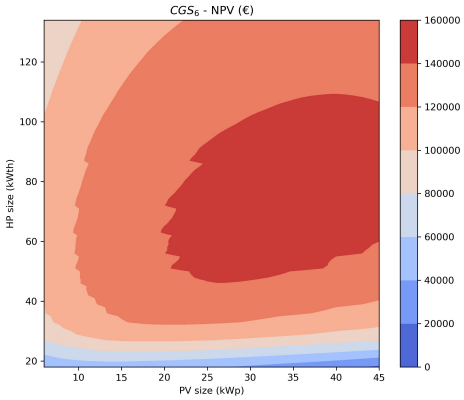
(c)



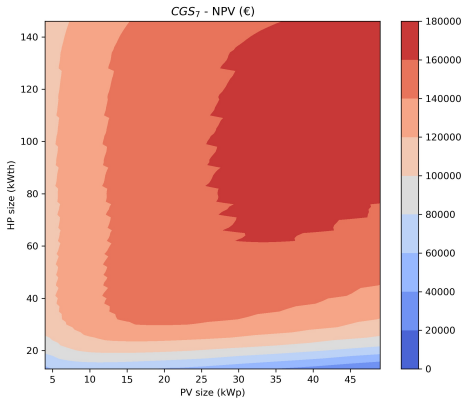
(d)



(e)

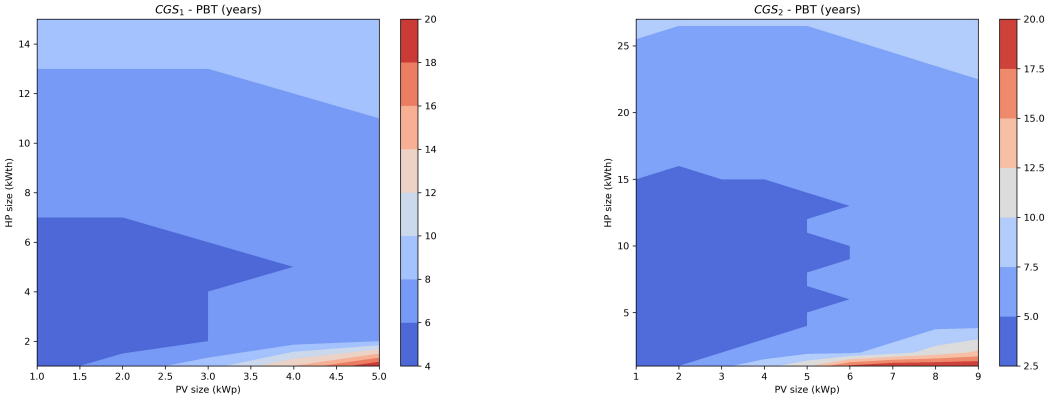


(f)



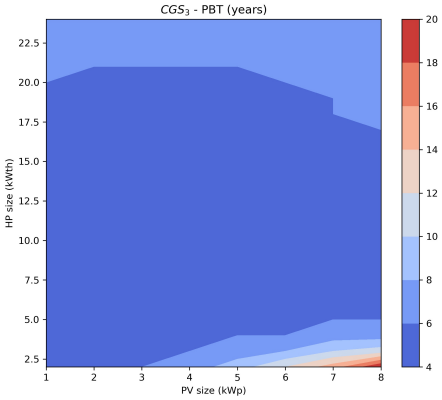
(g)

Figure 5.16: Net Present Value with control logic n° 3

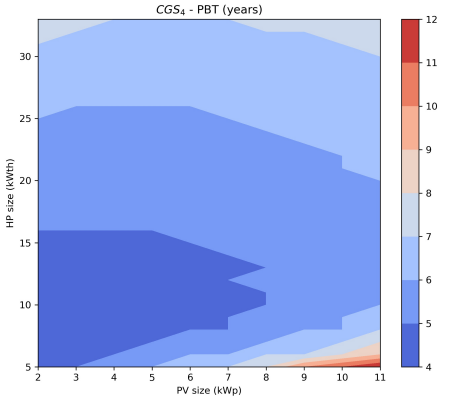


(a)

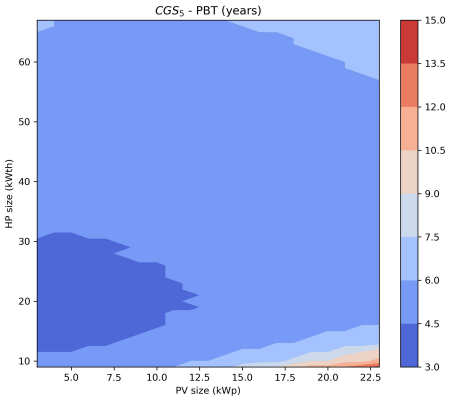
(b)



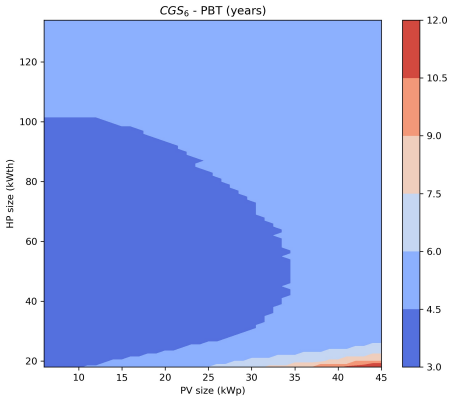
(c)



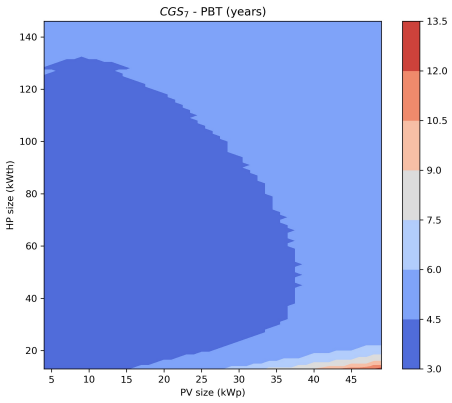
(d)



(e)

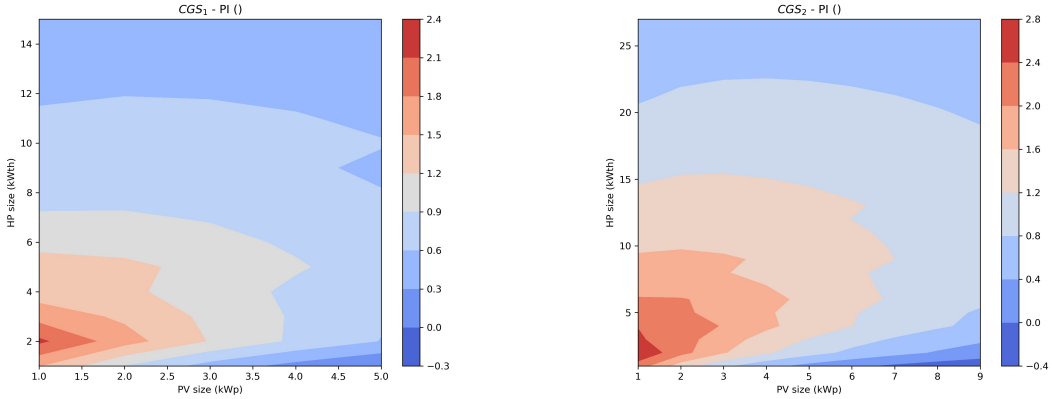


(f)



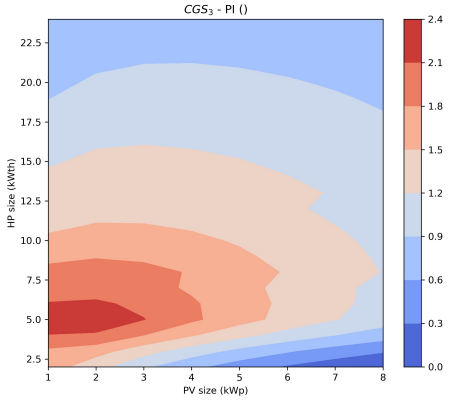
(g)

Figure 5.17: Pay Back Time with control logic n° 3

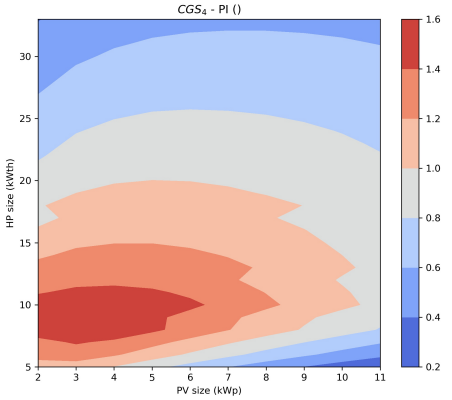


(a)

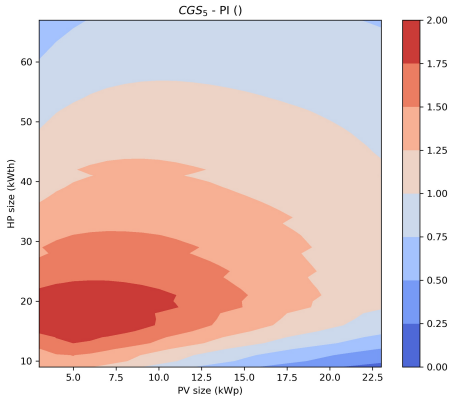
(b)



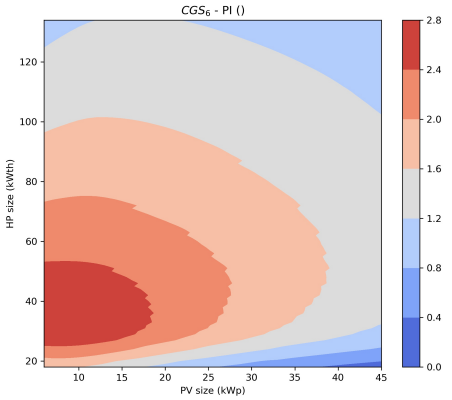
(c)



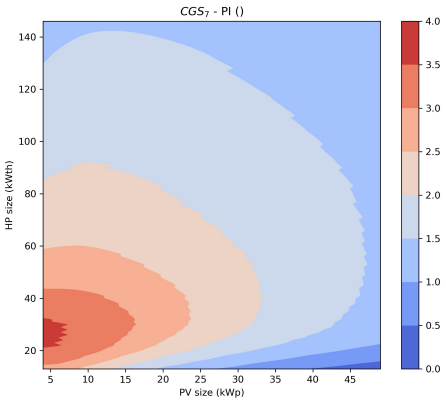
(d)



(e)



(f)



(g)

Figure 5.18: Profit indexes with control logic n° 3

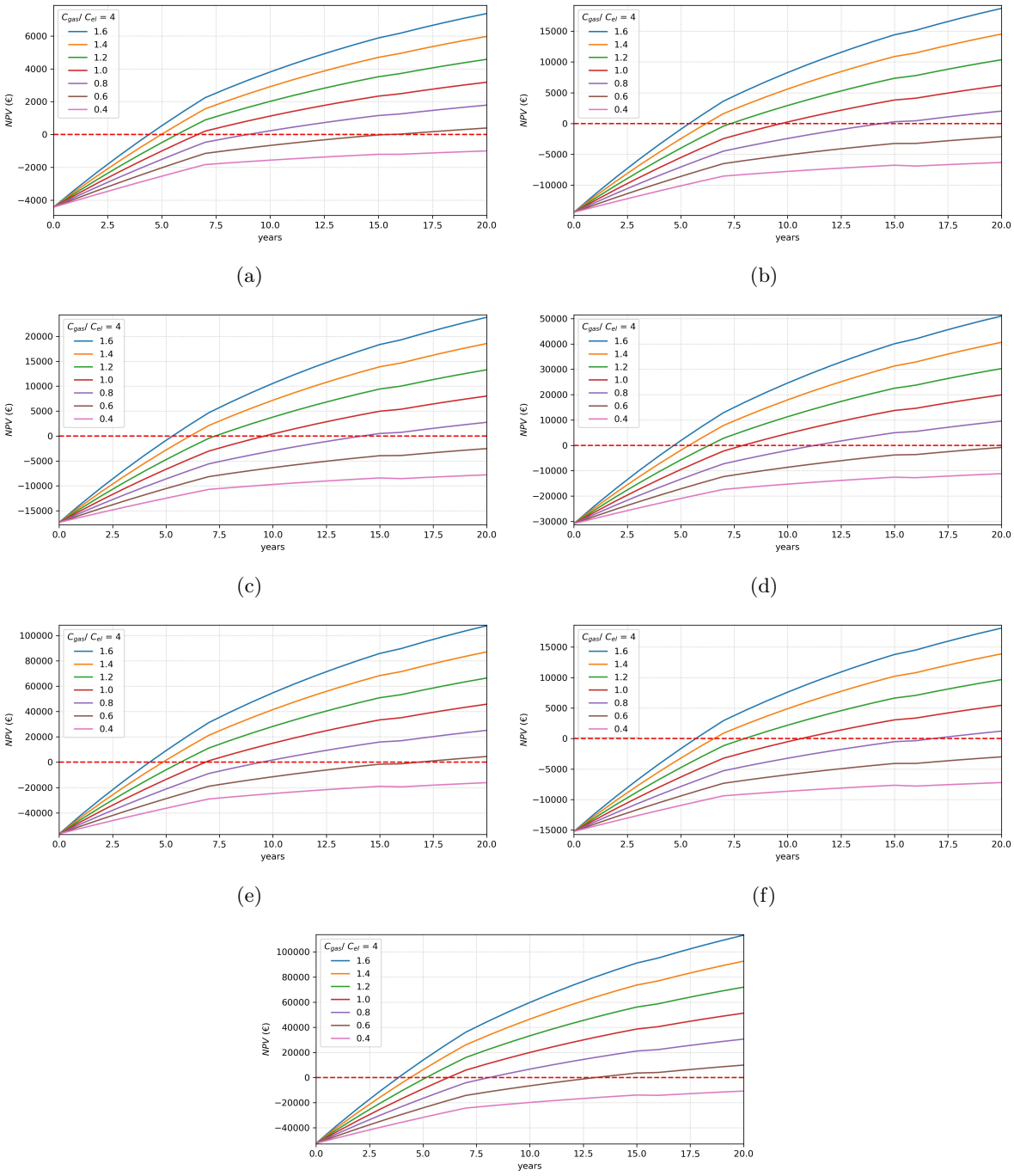
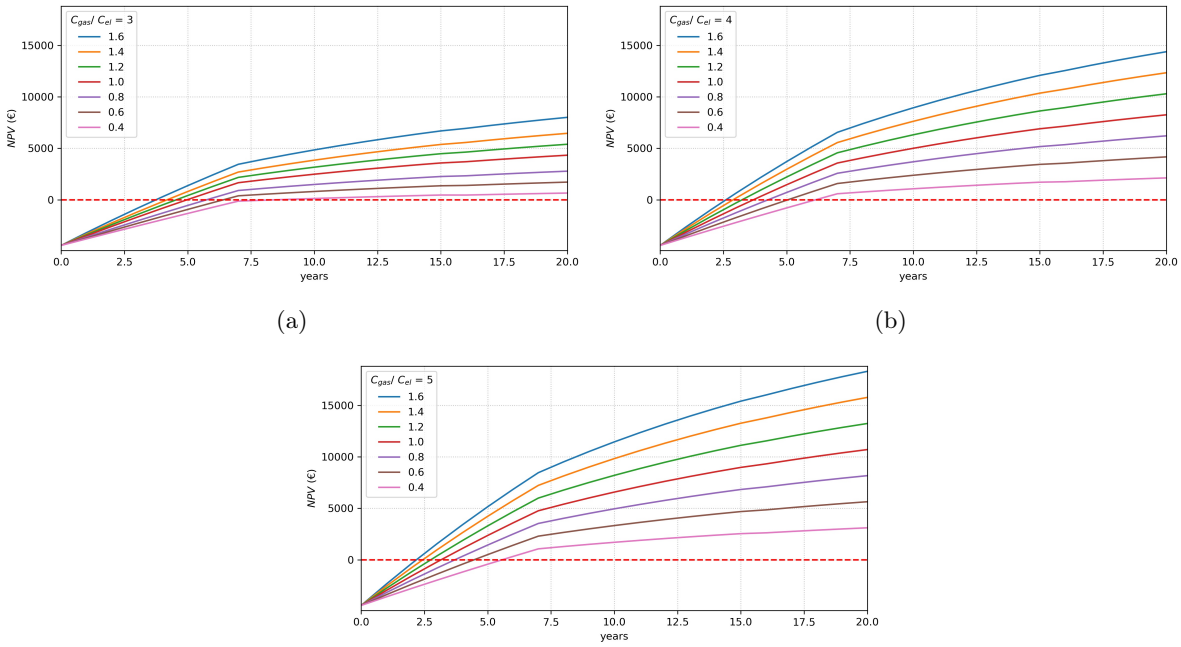


Figure 5.19: NPV sensitivty analysis with control logic n° 1 and n° 2.

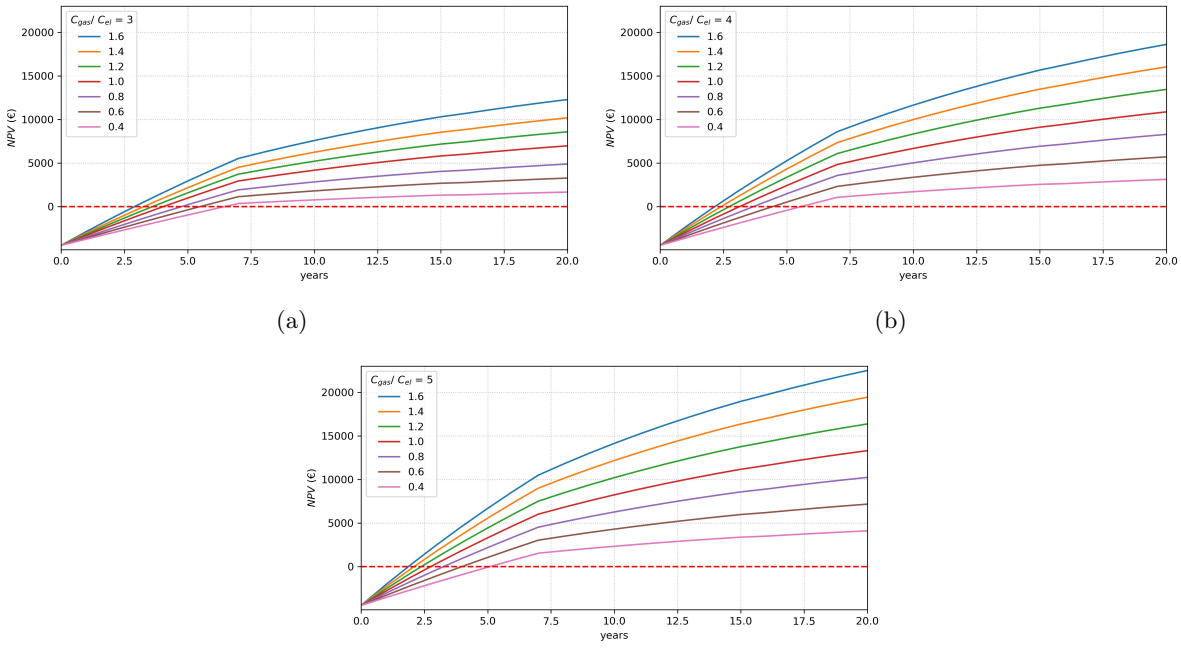


(a)

(b)

(c)

Figure 5.20: CGS_1 : NPV sensitivty analysis with control logic n° 3.

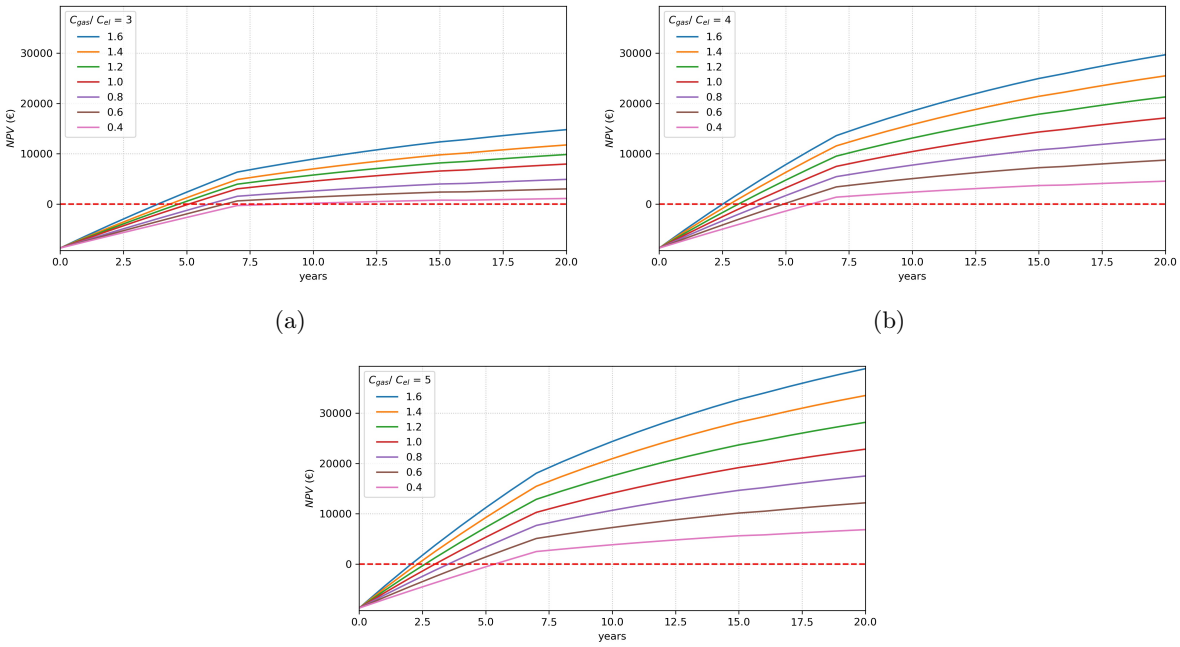


(a)

(b)

(c)

Figure 5.21: CGS_2 : NPV sensitivty analysis with control logic n° 3.

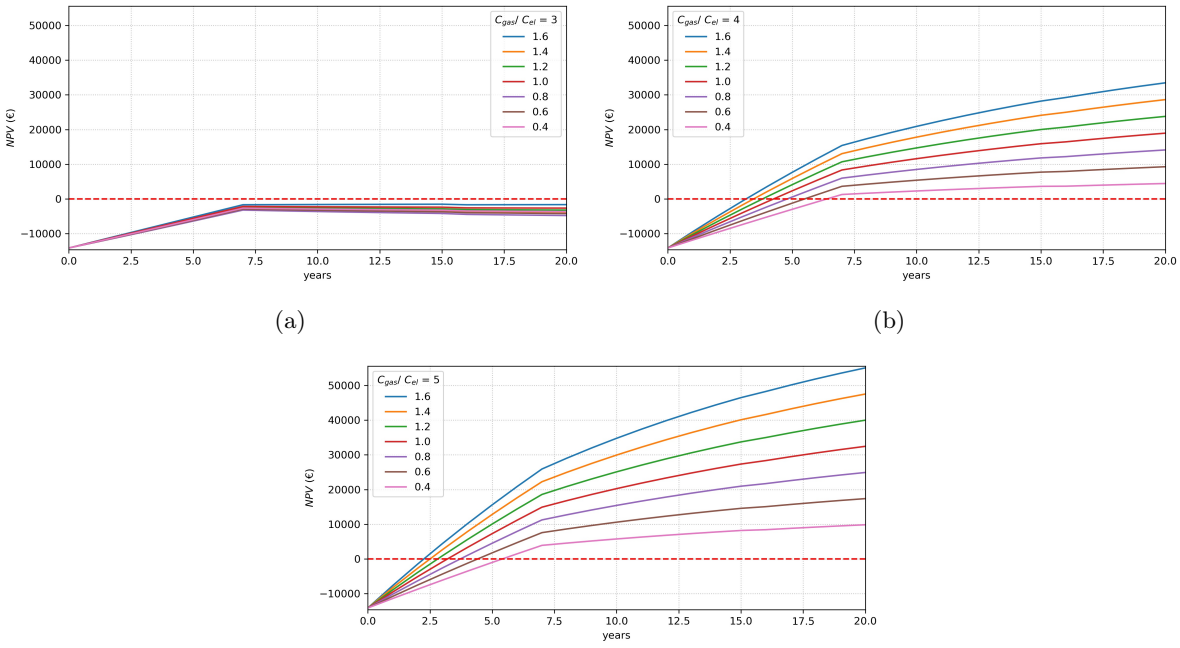


(a)

(b)

(c)

Figure 5.22: CGS_3 : NPV sensitivty analysis with control logic n° 3.

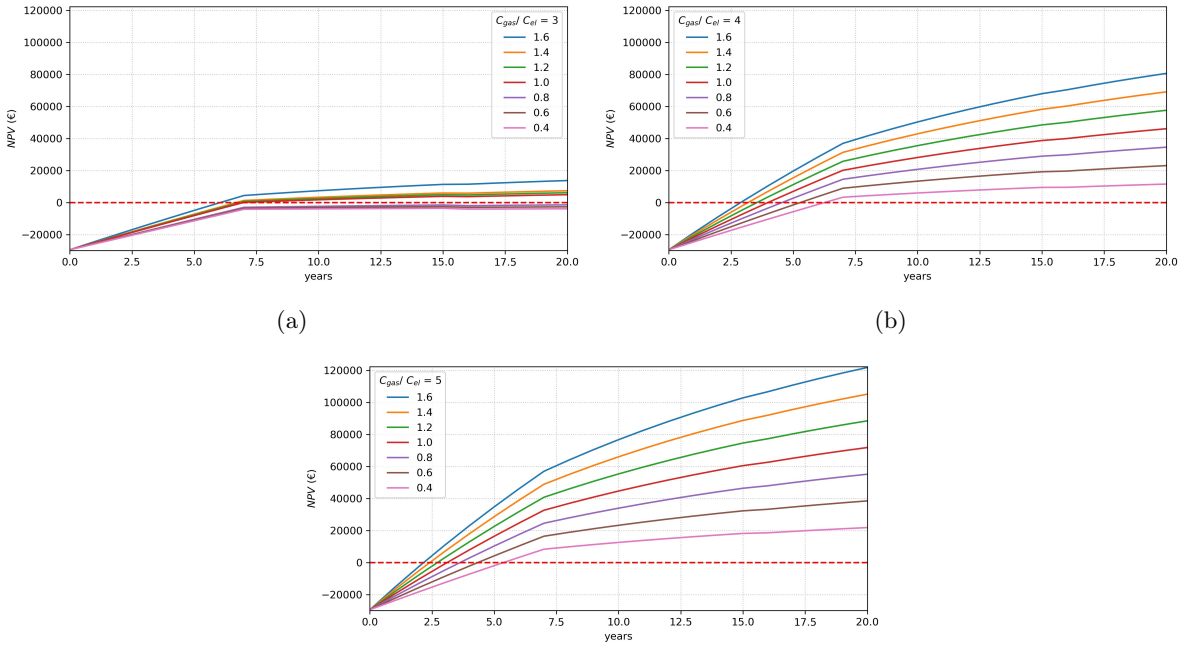


(a)

(b)

(c)

Figure 5.23: CGS_4 : NPV sensitivty analysis with control logic n° 3.

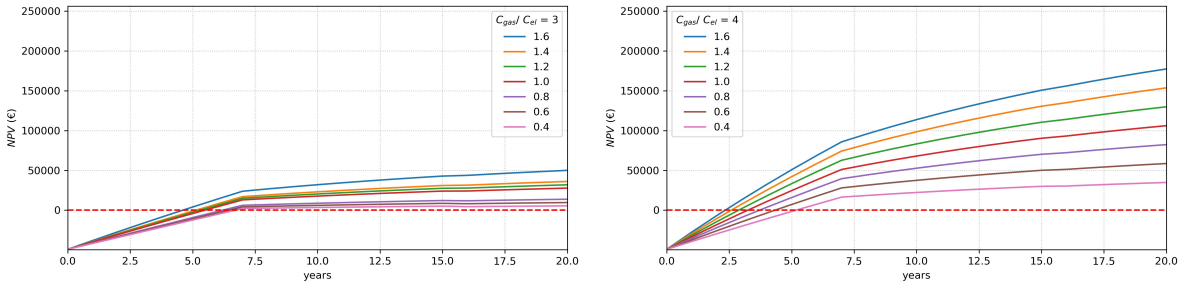


(a)

(b)

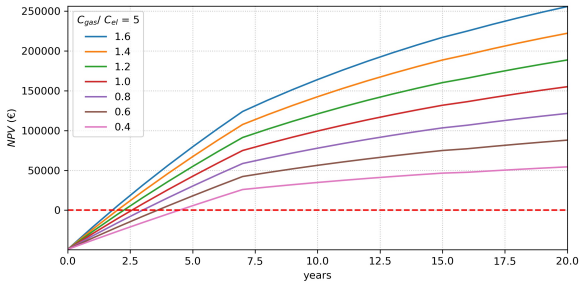
(c)

Figure 5.24: CGS_5 : NPV sensitivty analysis with control logic n° 3.



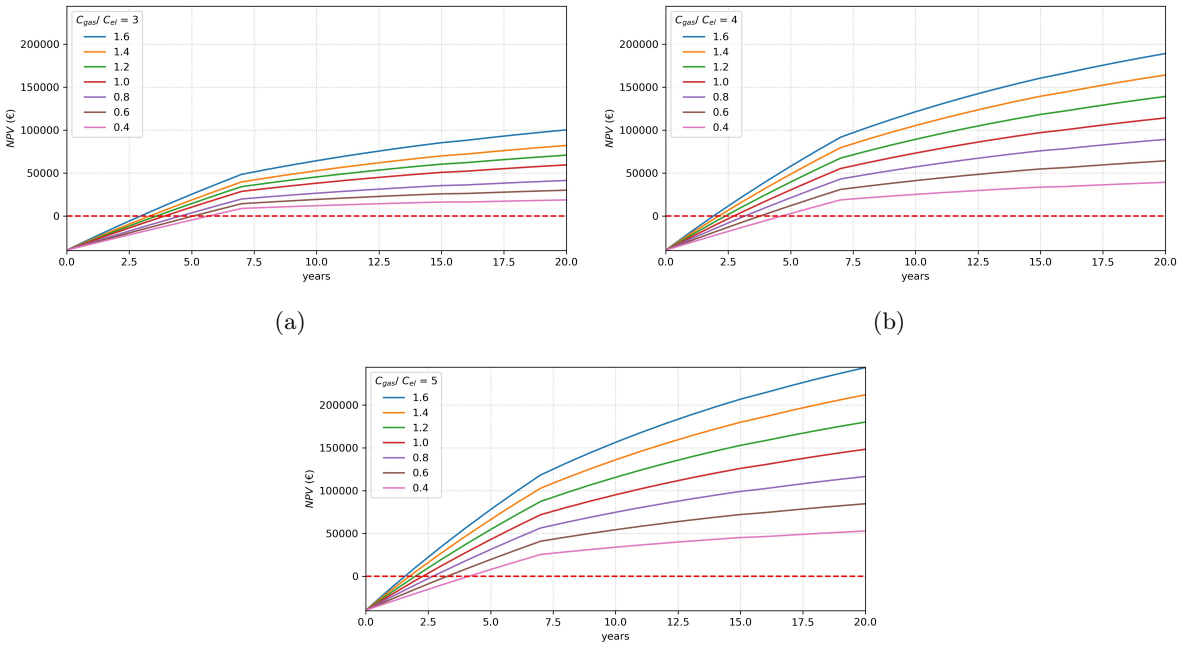
(a)

(b)



(c)

Figure 5.25: CGS_6 : NPV sensitivty analysis with control logic n° 3.



(a) (b)

(c)

Figure 5.26: CGS_7 : NPV sensitivty analysis with control logic n° 3.

5.3 Decarbonization analysis

In the last paragraph, we analyze the effects on the actual decarbonization of the preheating process, considering the choice of the techno-economic best solution for each control logic applied and for each plant at the same time. The first figure (5.27) shows the decarbonization that can be achieved by the optimal techno-economic solutions related to control logic 1 (off-grid) and 2 (on-grid with sales only): this is the same for both, as the second control logic only differs from the first in that the renewable surplus is not wasted, but sent to the grid. The plant that achieves the highest level of decarbonization is number 3, with a value of around 27/28 % of the total annual pre-heating energy, while the one that achieves the lowest decarbonization is number 1. This result is the combination of two factors: the coupling of the curve of each plant with the production curve of the photovoltaic and then the heat pump. In any case, the first result that we can highlight is that with these two control logics and in this given configuration of heat pump plus tank plus auxiliary boiler, for all the systems considered, it is possible to achieve a percentage of decarbonization of the preheating ranging from around 15% to a maximum of just under 30% of the total energy required in a year. The second figure, 5.28, on the other hand, shows how the inclusion of control logic 3, i.e. on-grid system with the possibility of purchasing from the grid, changes the situation compared to the two previous cases. While in the first two cases, the systems that benefited most from the inclusion of the heat pump were the small- to medium-sized systems within the cluster of systems analyzed, in the case of control logic number 3 the trend is reversed: the systems benefiting the most from the decarbonization are numbers 4, 5 and especially number 6. The greater decarbonization in the off-grid case is easily explained by the coupling of the preheat demand curve and the supply of energy from renewable sources. The C and F plants have the most favorable SSR and SCR factors in combination all. CGS_3 has a maximum SSR of 36% while maintaining a minimum SCR of 40%, while CGS_3 has a maximum SSR of 33% with a minimum SCR of 40%.

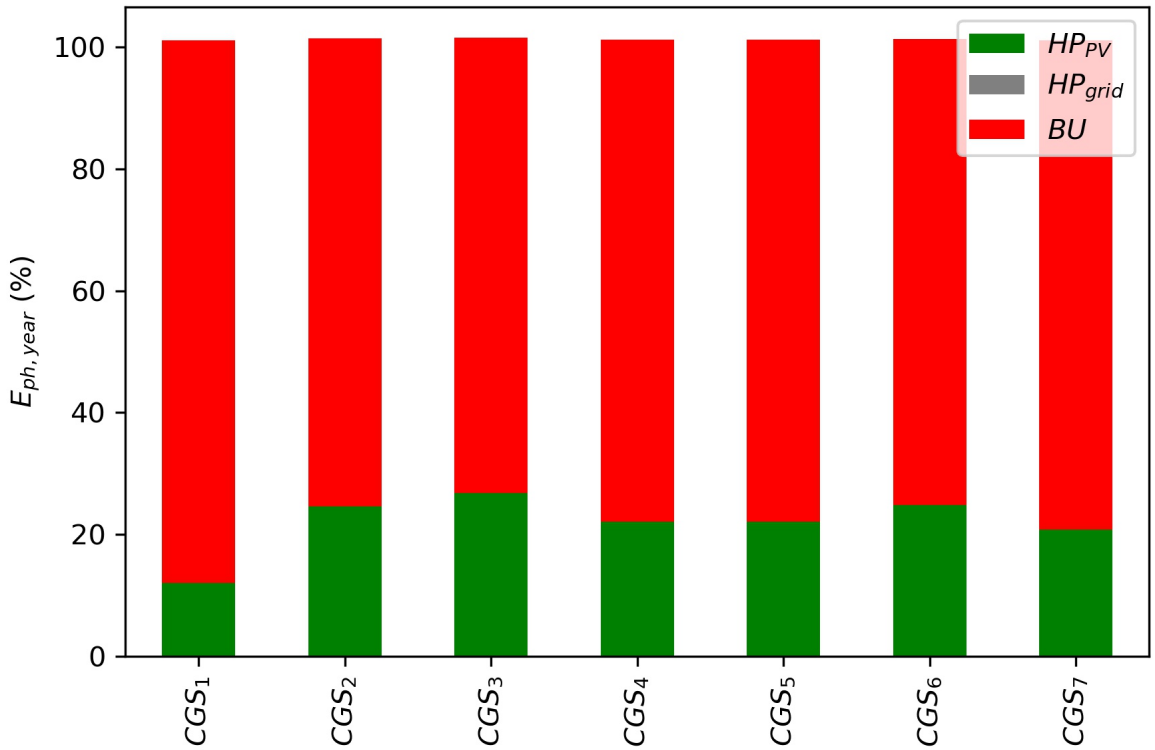


Figure 5.27: Decarbonization of gas preheating with control logics 1 and 2.

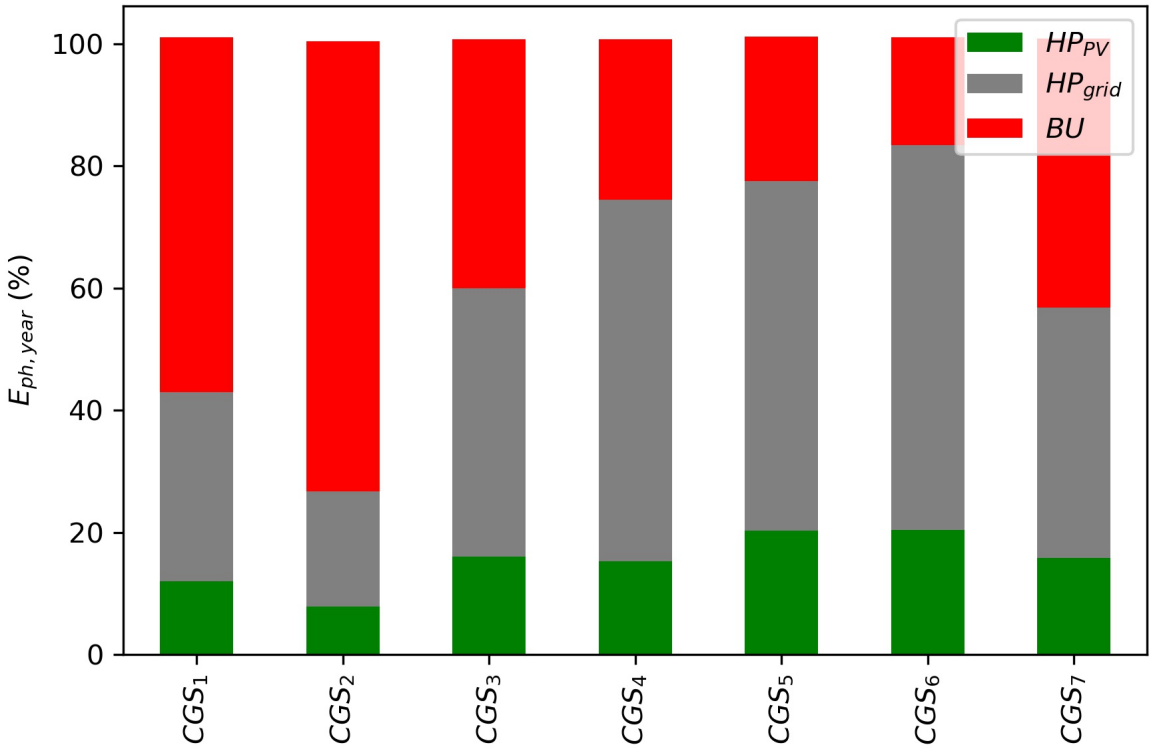


Figure 5.28: Decarbonization of gas preheating with control logic 3

Considering the second figure, it is possible to reduce the environmental impact of the preheating system in hybrid form (part fully decarbonized and part partially decarbonized) for the following percentages of total energy required annually: 42 %, 28 %, 60 %, 76 %, 79%, 84% and 57 % for plant from 1 to 7, respectively. Beyond what can be deduced from the optimum cases just analyzed and described in this section, other conclusions can be drawn. If the optimal cases took into account the simultaneous optimization of both the return on investment, thus the IP, and the payback time, a fundamental value for convincing DSOs to implement efficiency policies in a scenario of extremely volatile and uncertain energy prices, it is possible to obtain confidence bands in which to move around the regions of techno-economic optimal. This could help the DSO choose the best configuration depending on the logic being implemented and the class of plant that is to be decarbonized.

- For logic No. 1 and No. 2 the tolerance bands are not very wide, this is because the decoupling between the supply and demand curves, as often repeated in this discussion, means that beyond a certain value of HP and PV sizes the investment is never worthwhile. Specifically, the upper right corners (maximum HP and minimum PV) and lower left corners (minimum HP and maximum PV) will yield PBT values of 16 years or more and up to 20 years. It is not possible to recover more than 27 %, 36 %, 36 %, 30 %, 33 %, 33 %, 32 % with logics 1 or 2 for plants 1 to 7 respectively. In these cases then the most obvious thing is to seek the trade-off between maximum recovery and minimum expenditure, as is precisely obtained from the results of the techno-economic analysis.

- For logic No. 3, on the other hand, an additional consideration can be added, the possible energy recovery is theoretically total, so 100 % of the preheating energy can be provided by the HP, which, however, will be supplied on an annual basis by a mix of 100 % renewable energy and partially renewable energy. If one wanted to keep under 8 years as PBT, it would be possible to recover 100% of the preheat energy for all plants with the assumptions chosen for analysis.

5.4 Concluding remarks

Through this work, it was possible to analyze a series of innovative plant solutions that can be easily implemented by the DSO to decarbonize the pre-heating request of the CGS. The pair of optimal PV and HP sizes follows a generally similar trend for all 7 classes of systems analyzed. The techno-economic optimum was evaluated taking into account all three indices, namely NPV, PBT, and PI, but for a DSO it is essential to maximize the return on the investment and obtain results as soon as possible, to the detriment of a final profit superior. The techno-economic analysis gave this priority first to PI, then to PBT, and finally to NPV. The shape of the statistical distribution of the flow rate of the plants then influenced the actual percentage of final decarbonization. Moreover, the solutions were very dependent on the fluctuation of absolute prices and the ratio between the price of gas and that of electricity and it has not been possible to decarbonize 100% of the pre-heating without distancing from the optimum of the techno-economic analysis. The main outcomes of this first part of the work are listed below:

- The use of a super-simplified model, described in Chapter 2, allows for the analysis of possible decarbonization scenarios for any CGS-type plant with low computational effort.
- Each plant in the analyzed cluster presented a different cumulative preheat demand curve over the year, meaning different demand distributions within the plant's operating range. The percentage of preheat energy was calculated as about 0.2 % of the energy flowing through the CGS in a year.
- The three control logics differ in that they can interact with the network in different ways. Only with logic No. 3 is it possible to use the HP all year round, but it is never possible to exceed an SSR of about 36 % and maintain an SCR greater than 48 % regardless of which logic is active.
- The techno-economic analysis shows that investments in low to medium sizes of PV and HP are the most convenient in terms of PBT and PI for any plant and for any control logic implemented. Considering the NPV index, the choice should fall on plants with low-medium sizes of HP and medium-high sizes of PV, especially for logic 2 and 3, which is when surplus electricity can be sold. Tecnho-economic optimum for each plant with each control logic is obtained.
- The sensitivity analysis of the techno-economic optimum clarifies the dependence of the goodness of investment on the gas price and the gas/electricity price ratio. The investment maintains its profitability over time with definitely sustained gas prices and at least equal or higher 0.6-0.8 €/Smc and with increasing gas-electricity prices ratios, that is, toward scenarios in which gas and electricity prices are decoupled.
- From the decarbonization analysis, it emerges that it is not possible to recover all the pre-heating energy with the first two logics, but only with the last logic (totally on-grid). However, it is never possible to completely decarbonize the system, even in cases where energy from the grid is used and the size of HP is such as to cover 100% of the annual pre-heating gas requirement. The maximum is approximately 28% with only 100% decarbonized energy or approximately 84.7% with a mix of 100% renewable energy and energy from the grid, out of the total energy required by preheating in the CGS.

Chapter 6

Results: energy monitoring of CGS with machine learning models

In this chapter, the method developed for data-driven modeling of preheat consumption in CGSs is presented and then its application to two different case studies is described. The chapter is developed as follows: first, the results of the feature selection and engineering process are presented, then the results of the machine learning models in terms of performance are discussed, and finally, the models are applied to the two real cases using the CuSUm technique to evaluate the energy performance of the plants.

6.1 Features selection and engineering

When approaching the modeling of data-driven regression algorithms, the first step to be tackled concerns the correlation analysis that is present between the output variable, the consumption target in this case, and the input variables, the features related to the sensors installed in the NG's pressure-reduction plants. Figures 6.1 and 6.2 show the Pearson correlation matrix for all the raw features and target contained in the two datasets of CGS_A and CGS_B, respectively. Considering the basic sensors and not those related to the preheating system (set points, preheating water temperatures, etc.), the natural gas flow rate Q_{gas} processed by CGS is highly correlated with gas consumption, as expected, with very high correlation coefficient values (above 0.9 and specifically 0.95 and 0.92) and also the ambient air temperature T_{env}, which gives an indication of the seasonality and consequently the gas consumption of downstream consumers. Environmental air temperature has a high and negative correlation value (less than -0.7) because as the temperature increases (summer), consumption decreases and vice versa (winter). The temperature of the pre-heating water, i.e. the hot water supply T_{wat}, also shows high values of the correlation coefficient (around 0.7) concerning consumption and for the CGS_A system. The correlation between CGS_B consumption and T_{wat} is even higher due to the fact the the second plant has a central management system that adapts the water temperature set point accordingly to the gas flow rate. The average correlated sensors include the gas inlet pressure P_{in} (absolute values between 0.6 and 0.4), as a more pressurized gas will require a greater temperature jump to compensate for a stronger Joule-Thompson effect, and the outlet gas temperature T_{out}, as outlet conditions are strongly influenced by the preheating consumption: on summer the system is capable to preheat better a small amount of volumetric flow rate and the set point is set at a higher value equal to 10°C, while in the winter the gas outlet set point is lowered to 8°C. To summarise, all the inputs related to gas inlet conditions (flow rate, temperature, and pressure) and environmental conditions are among the main quantities by order

of correlation between standard sensors. Another sensor that gives us information on the seasonality and operating conditions of the CGS is the gas outlet set point temperature, which is varied once a year according to the type of 'thermal' season, i.e. summer (second half of spring, summer, and first half of autumn) and winter (second half of autumn, winter and first half of spring). The solar radiation G and the outlet pressure P_{out} are two variables that are correlated on average (between 0.4 and 0.2), as the first one provides additional and complementary information on seasonality and the second one is less important for the gas preheating consumption since it strictly follows the outlet pressure set point (downstream network operating pressure). As far as the correlations between characteristics are concerned: irradiation G is on average correlated with the ambient temperature (0.53) as expected from their seasonal relationship, and the inlet pressure also has a good correlation with the total gas flow rate (around 0.55) as the systems work proportionally to the increase in flow rate with the increase in upstream pressure due to the transport network. On the other hand, the outlet pressure is poorly (< 0.2) correlated with the other variables, especially considering the CGS_B plant.

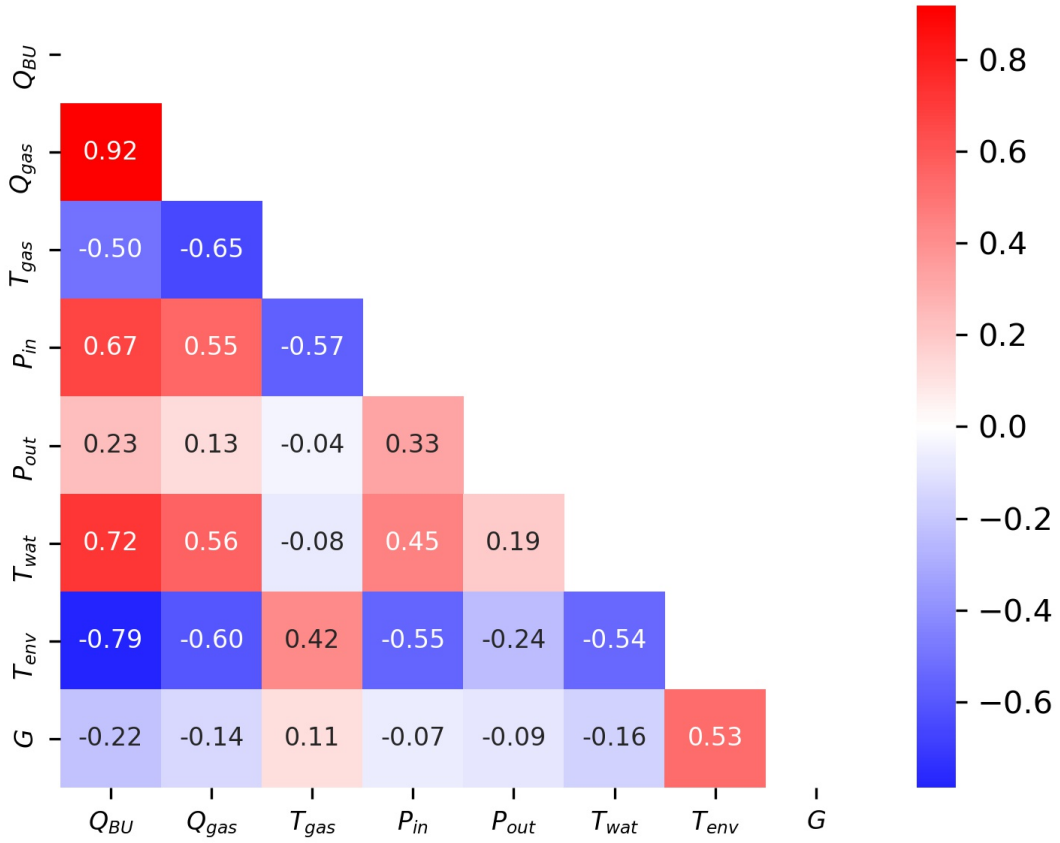


Figure 6.1: Pearson correlation matrix: CGS_A dataset

A further consideration regarding the variables present in the heating plant management system, active only in the CGS_B plant: there are the boiler ignition minute counters (from 0 to 60), the boiler ID BU_{ID} which is a counter providing the operating condition of the heating plant (0 boilers on, 1 boiler on and which, 2 boilers on at the same time) and the heat pump timer (from 0 to 60). The first two are highly correlated, but are very prone to data leakage problem as they insert a forcing of information in the features that replicate the target and the risk of not having the features in the inference process would make the model very inaccurate.. In statistics and machine learning, leakage (also known as data leakage or target leakage) is the use of information in the model training process which would not be expected to be available at prediction time, causing the predictive scores (metrics) to overestimate the model's utility

when run in a production environment For these reason they have been preliminary rejected from the features candidates.

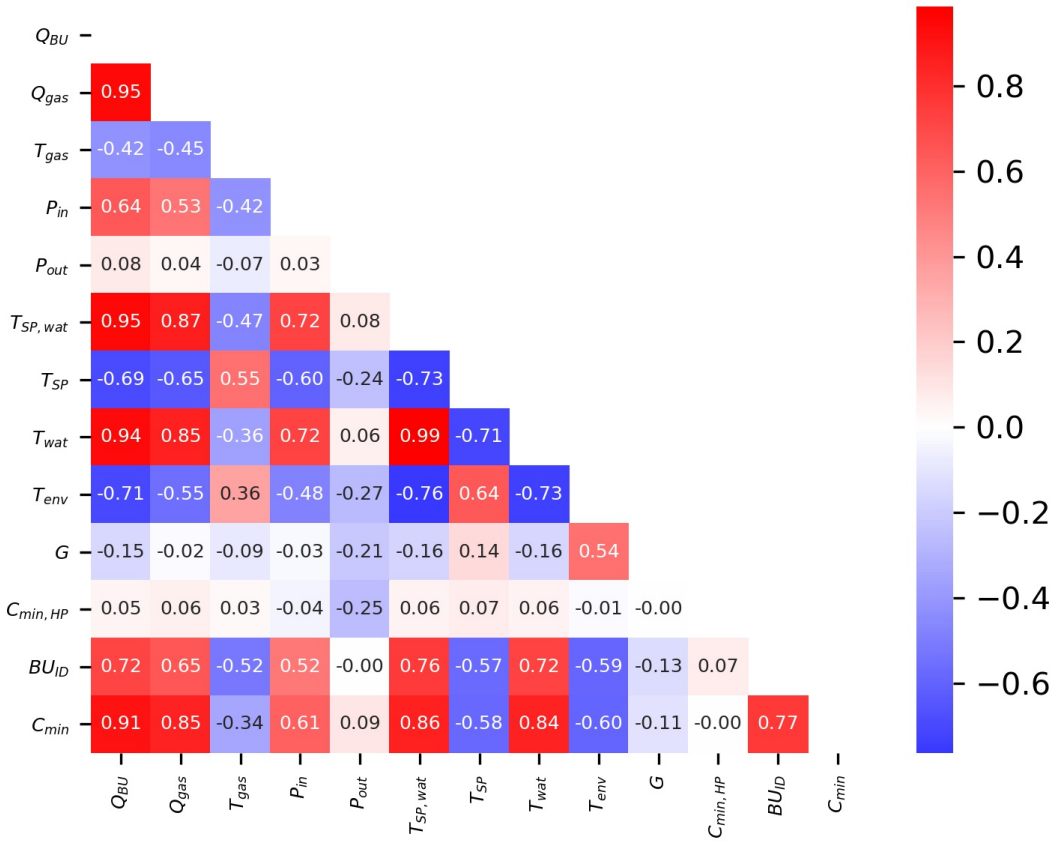


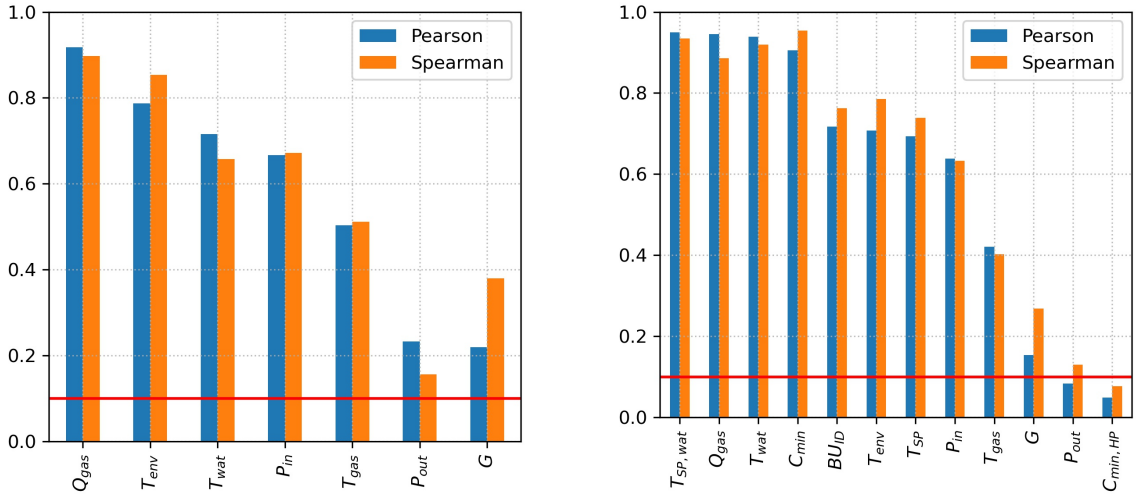
Figure 6.2: Pearson correlation matrix: CGS_B dataset

Figures 6.3(a) and 6.3(b) show the barplot of the correlation absolute values, from 0 to 1, of the two correlation indexes, Pearson and Spearman, between the target, QBU, and all of the other dataset variables, the possibile candidate features, for each plant. In these graphs, instead of also analysing the correlation between all features, we only analyse both the linear (Pearson) and monotonic (Spearman) correlation between the features and the target Q_{BU}: it is interesting to note that variables often have a Spearman's coefficient correlation value greater than Pearson's, which means that there is often a non-linear correlation between targets and feature candidates.

6.1.1 Features engineering

Feature engineering is a process whereby new information and thus new quantities are extracted from the available raw data. The following engineered features were extrapolated for this case:

- $T_{gas,in}$ = inlet gas temperature, obtained with equation 2.23 as a function of the ambient temperature using established models from the literature and described in the previous chapter.
- Month = month of the year counter (assuming values from 1 to 12)
- Time of the day in terms of sinusoidal functions ($Sin(H\pi/12 + \phi)$, $Cos(H\pi/12 + \phi)$. The two sinusoidal fuctions are correlation-optimized with respect to the target. Twenty-four values are tested, one per hour, in order to find the sine function with the correct shift value ϕ that gives the highest possible Pearson and Spearman coeficients together. The two figures show the trends of the



(a) (b)

Figure 6.3: Correlation barplots between target and raw features: CGS_A (a) and CGS_B

sinusoidal functions and how they are trying in some way to replicate the trend of the natural gas consumption curve. In the case of CGS_B the two functions placed the extremes (1 and -1) near the peaks, while the opposite happens for CGS_A which obtained sines and cosines in the presence of the valleys.

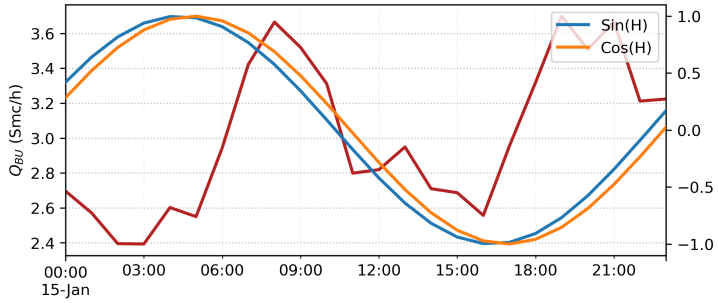


Figure 6.4: Optimized sinusoidal engineered features for CGS_A

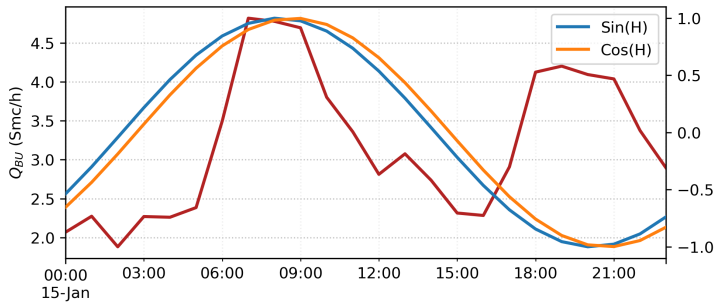


Figure 6.5: Optimized sinusoidal engineered features for CGS_B

- Weekday = counter for the day of the week (assuming values from 1 to 7)
- Daytype = counter for the type of the day, such as workday, pre-festivity or festivity (assuming values from 1 to 3)

Time expressed in sinusoidal form makes it possible to eliminate the discontinuity of values that occurs in the transition from one day to another, that is, in the transition from time 23:00 to time 00:00. In

addition, it is important to use both in combination since their sum expresses the completeness of the hour in sinusoidal form.

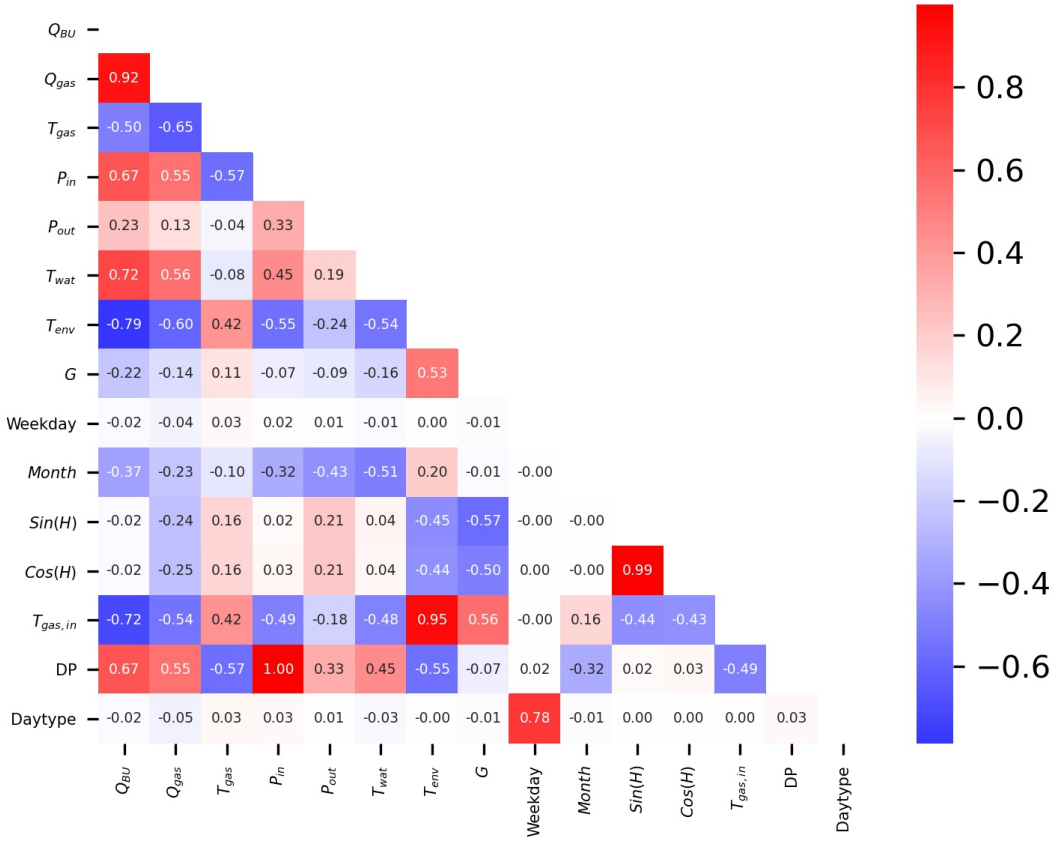


Figure 6.6: Correlation matrix with engineered features: CGS_A.

Once the engineered features are chosen and created, a complete starting dataset for each thermal plant is obtained. Figure 6.6 shows the Pearson correlation matrix for the CGS_A and figure 6.7 shows the Pearson correlation matrix for the CGS_A and CGS_B.

Features selection is the phase in which the inputs to be fed to the training of the various machine learning models are finally chosen. for this application, the simple but computationally inexpensive method known as the filter method is chosen, which is a method based on the correlation between variables and which establishes a threshold of acceptability of the correlation value between the target variable and all other input variables. The filtering threshold is chosen to be 0.1 for both correlation coefficients used in this work. Figure 6.8 shows the results of the filtering process with two barplots, one for each plant. If both Spearman and Pearson coefficients are less than 0.1 for the target and candidate input quantity tuple, then the candidate variable is discarded and not used as input to the model in the subsequent training phase. For CGS_A, the gas outlet pressure, the time-of-day values in the sinusoidal form (Cos(H) and Sin(H)), the type of day (weekday, holiday, etc.), and the day of the week are discarded. On the other hand, for the CGS_B, the Cos(H) has a Spearman correlation value with the target greater than 0.1 and thus is retained. It can be seen from the figure that there is substantial similarity in the values of the correlation coefficients between the target and the candidates, for both plants used as case studies. In addition, all those variables that are themselves other targets or targets analogous to the one of interest for this task are discarded: the hot water outlet temperature and the gas outlet temperature (not the set point, but the actual temperature) are discarded because the first is almost perfectly correlated with the set point of its own magnitude, as is rightly expected, and the other because it is physically the result of

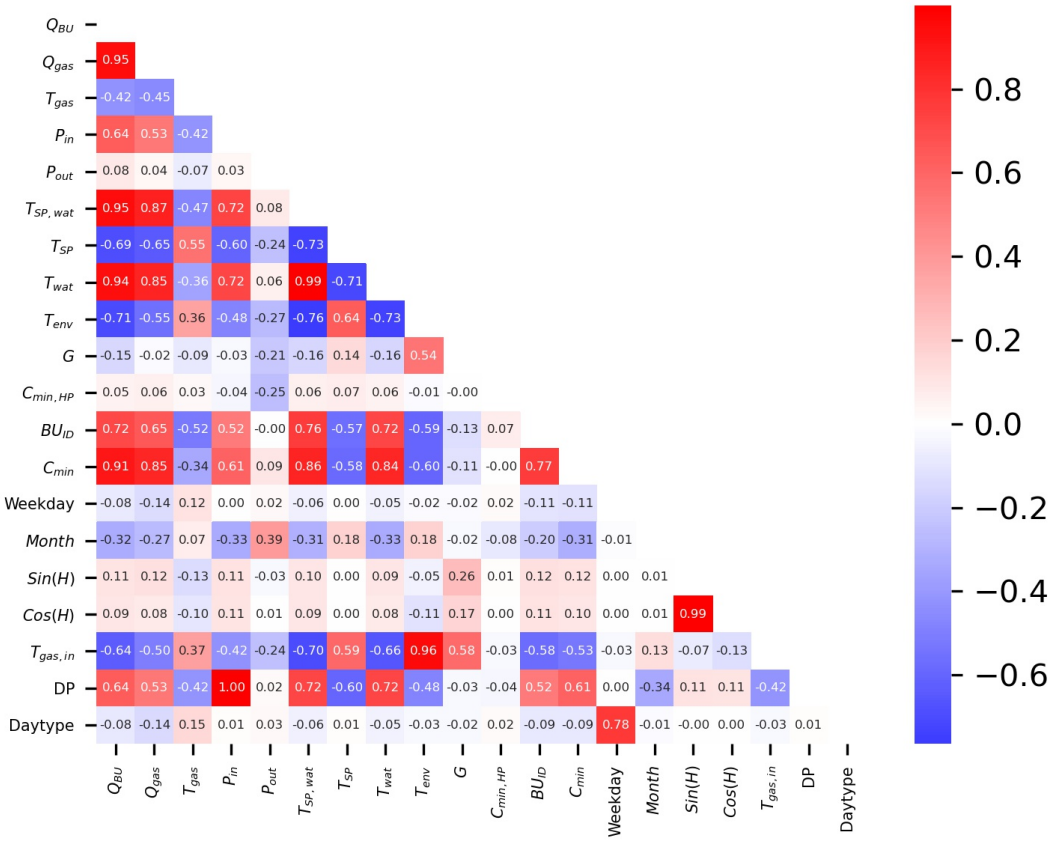
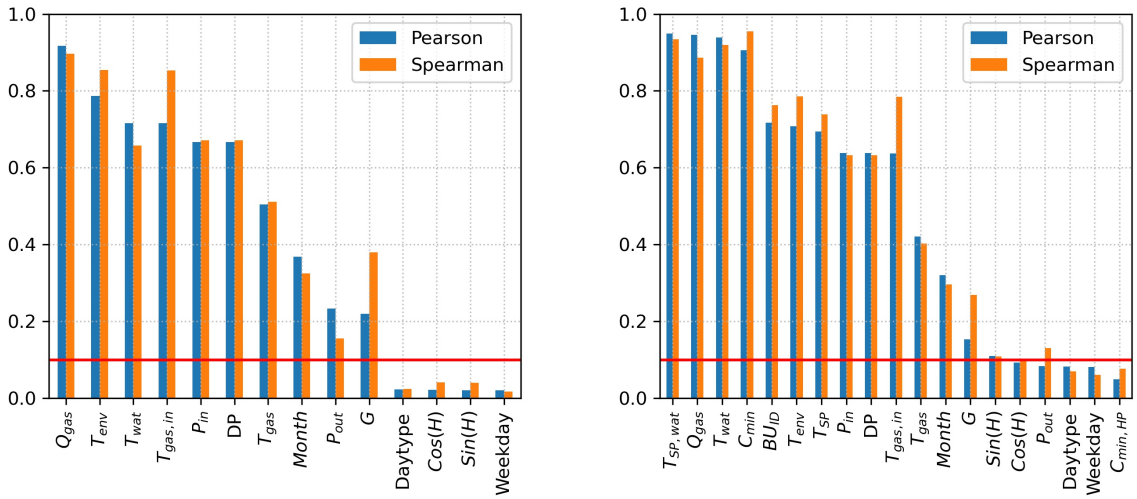


Figure 6.7: Correlation matrix with engineered features: CGS_B.



(a) (b)

Figure 6.8: Correlation barplots between target and engineered features: CGS_A (a) and CGS_B

consumption. The third variable eliminated is the DP pressure drop (i.e., the difference between P_{in} and P_{out}) since it adds no information compared to the two quantities taken individually. List of variables accepted and eliminated for each cabin after the features selection and engineering phase:

- CGS_A

  Features accepted: Q_{gas} , P_{in} , P_{out} , $T_{SP,wat}$, T_{SP}, T_{env} , G, $Month$, $T_{gas,in}$

  Features rejected: T_{wat} , T_{gas}, $Daytype$, $Weekday$, $Sin(H)$, $Cos(H)$

- CGS_B

  Features accepted: Q_{gas} , P_{in} , P_{out} , $T_{SP,wat}$, T_{SP}, T_{env} , G, $Month$, $Sin(H)$, $Cos(H)$, $T_{gas,in}$

  Features rejected: T_{wat} , T_{gas}, C_{min}, $C_{min,HP}$, $Daytype$, $Weekday$

6.2 MLR models deployment

In this section, the results of the MLR models for CGS_A and CGS_B plants are presented. The MLR model is trained using a dataset relating to the complete operation of the CGS for one year, specifically from April 2021 to April 2022 for the CGS_A and the CGS_B. The dataset is divided into two parts, 80% for the actual training of the model and 20% for testing the model. For both plants, the procedure was followed to create a baseline model of MLR, which was then compared with the final MLR model. The difference between the former and the latter is that only the raw features that can be directly attributed to preheating consumption are used (see Chapter 3 equation 2.13). In this way, the impact of the addition of engineered features on the predictive performance of the models can be quantified. Equation 6.1 shows the shape of the baseline MLR model. Equation (6.2) shows the structure of the MLR model and Table 6.1 shows the values of the multiplicative coefficients for each feature for plant CGS_A model. The c_0 coefficient is an independent term in the linear model of scikit-learn called intercept. Equation (6.3) shows the structure of the MLR model and Table 6.2 the values of the multiplicative coefficients for each feature for plant CGS_B. In the basic MLR models, for all plants the following variables are chosen as inputs: the flow rate of flowing gas Q_{gas}, inlet pressure P_{in}, outlet pressure P_{out}, and external environmental conditions (T_{env} and G). This makes it possible to obtain a total "black-box" model, that is, without having information of what happens inside the plant, but only knowing measurements of what happens directly upstream (Q_{gas}, P_{in}), downstream (Q_{gas}, P_{out}) and outside the plant (Tenv and G).

$$Q_{BU} = c_0 + c_1 \times Q_{gas} + c_2 \times P_{in} + c_3 \times P_{out} + c_4 \times T_{env} + c_5 \times G + \varepsilon \tag{6.1}$$

$$\begin{aligned} Q_{BU} = c_0 + c_1 \times Q_{gas} + c_2 \times P_{in} + c_3 \times P_{out} + c_4 \times T_{env} + \\ c_5 \times G + c_6 \times Month + c_7 \times T_{gas,in} + \varepsilon \end{aligned} \tag{6.2}$$

$$\begin{aligned} Q_{BU} = c_0 + c_1 \times Q_{gas} + c_2 \times P_{in} + c_3 \times P_{out} + c_4 \times T_{SP,w} + c_5 \times T_{SP} + c_6 \times T_{env} + \\ c_7 \times G + c_8 \times Month + c_9 \times Sin(H) + c_{10} \times Cos(H) + c_{11} \times T_{gas,in} + \varepsilon \end{aligned} \tag{6.3}$$

Table 6.1: CGS_A : Final MLR model coefficients.

Coefficient	Value	Coefficient	Value	Coefficient	Value
c_0	2.344785	c_3	-0.000605	c_6	-0.0371793
c_1	0.002270	c_4	-0.042019	c_7	0.0127065
c_2	0.014580	c_5	0.0001643	ε	278.03751

Table 6.3 shows the value of the different scoring metrics comparison for the MLR baseline model and the MLR final model for the first plant, namely CGS_A. The MAPE values achieved by the MLR model

Table 6.2: CGS_B : Final MLR model coefficients.

Coefficient	Value	Coefficient	Value	Coefficient	Value
c_0	-4.056984	c_5	0.040171	c_{10}	0.446430
c_1	0.001718	c_6	-0.020291	c_{11}	0.051859
c_2	0.005871	c_7	0.000163	ε	367.453
c_3	0.000469	c_8	-0.015831		
c_4	0.049998	c_9	-0.425531		

are discrete and are around 16% -15% for the basic and final models, respectively. The MAE results tend to reflect those of MAPE since MAE fluctuates between 0 and 1 as does MAPE between 0 %and 100 %. The only difference is that MAE goes one step further, by adding in the actual value division to convert it to a percentage. The MSE values are also good and equal to 0.166628 and 0.150308. However, what makes us understand the actual performance of the model are the parameters called RMSE, as they are dimensional and measured with Smc/h units. The percentage RMSE also is obtained by dividing the RMSE by the maximum value of the target, i.e., y = Q_{BU}, and thus they make us understand respectively that the model predicts the target with an absolute error of about 0.2 Smc/h and percentage of about 5%. The values of R^2 are very high and satisfactory, above 0.94 for both models. The graphs of the error distribution 6.9a and the predicted vs. actual scatter plot 6.10 show the tendency of MLR models not to be perfectly generalized for the case study: we see an imbalance of positive outliers (average overestimation of the value by the model for plant A).

Table 6.3: CGS_A MLR models comparison for test dataset: base vs final.

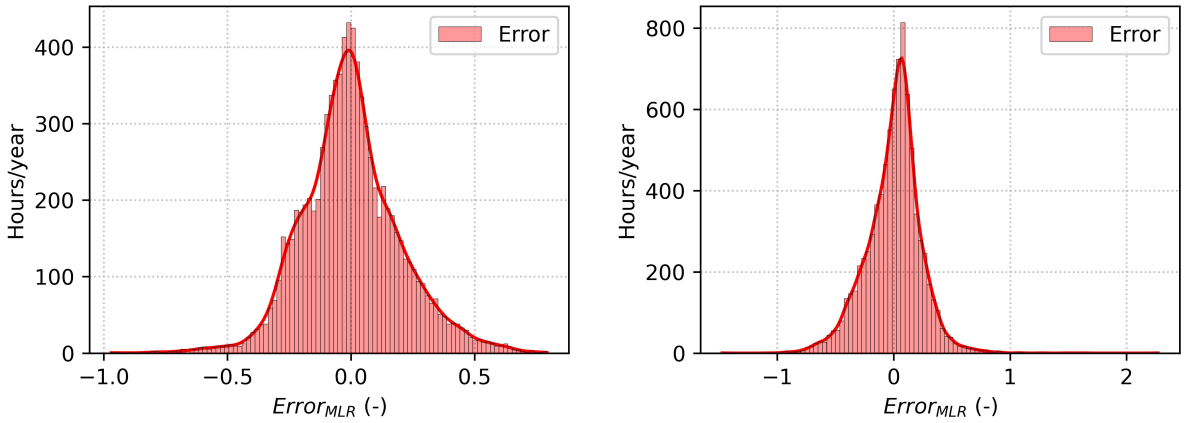
	Unit	Base	Final	Final vs Base
MAPE	[%]	16.33412	15.83054	-3.08%
MSE	$[Smc/h^2]$	0.050089	0.039532	-21.08%
MAE	[Smc/h]	0.166628	0.150308	-9.79%
RMSE	[Smc/h]	0.223806	0.198826	-11.16%
RMSE%	[%]	5.557637	4.937314	-11.16%
R^2	[-]	0.947101	0.958251	1.18%

Table 6.4 shows the value of the different scoring metrics comparing them for the MLR baseline model and the MLR final model for the second plant, namely CGS_B. The MAPE values achieved by the MLR model are higher than for the first plant and are around 38% and 27% for the basic and final models, respectively. This is beacause the second plant MLR model is affected by range of overestimating values and the MAPE is much more sensitive and asymmetric and it puts a heavier penalty on negative errors (when forecasts are higher than actuals) than on positive errors. The MSE values are on the other hand good and equal to 0.090154 and 0.054624 and better if compared to the first plant model. The RMSE and the percentage RMSE shows respectively that the model predicts the target with an absolute error of about 0.3 and 0.23 Smc/h and percentage of about 5% - 3%. The values of R^2 are very high and

satisfactory, above 0.95 for both models (0.7 for the model with engineered features included). At the same time, both MLR models achieved very high regression performance and were therefore retained as candidates for the consumption baseline of the two plants. The graphs of the error 6.9b and predicted vs. actual 6.11 values also confirm for us the tendency of MLR models not to be perfectly generalized for the case study: we see an imbalance of positive outliers (average overestimation of the value by the model for plant B).

Table 6.4: CGS_B MLR models comparison for test dataset: base vs final.

	Unit	Base	Final	Final vs Base
MAPE	[%]	38.50541	27.94834	-27.42%
MSE	$[Smc/h^2]$	0.090154	0.054624	-39.41%
MAE	[-]	0.221811	0.176396	-20.47%
RMSE	[Smc/h]	0.300256	0.233718	-22.16%
RMSE%	[%]	5.078755	3.873353	-23.73%
R^2	[-]	0.950997	0.970904	2.09%



(a) (b)

Figure 6.9: (a) CGS_A MLR errors distribution vs (b) CGS_B MLR errors distribution

The two scatter plots are plotted by dividing the dataset into training (blue) and testing (green). This is used to see whether the model is overfitting or underfitting by comparing the results of model predictions by graphing. In the tables just described, the value of testing is rightly reposted, but it is also important to observe the performance on training and testing simultaneously. For MLRs, there is no validation phase. As can be seen from the two figures, both models achieved an excellent level of generalization, with the test and training points statistically evenly distributed and equispaced concerning the best-fit line in red. For the first cabin, a greater dispersion of the predictions is observed for medium-high values of consumption, while for the second plant, a deviation of the scatter trend from the best-fit line is observed for medium-low values of consumption, i.e., when preheating consumption is limited and the effect of losses on the total is more incisive in percentage terms.

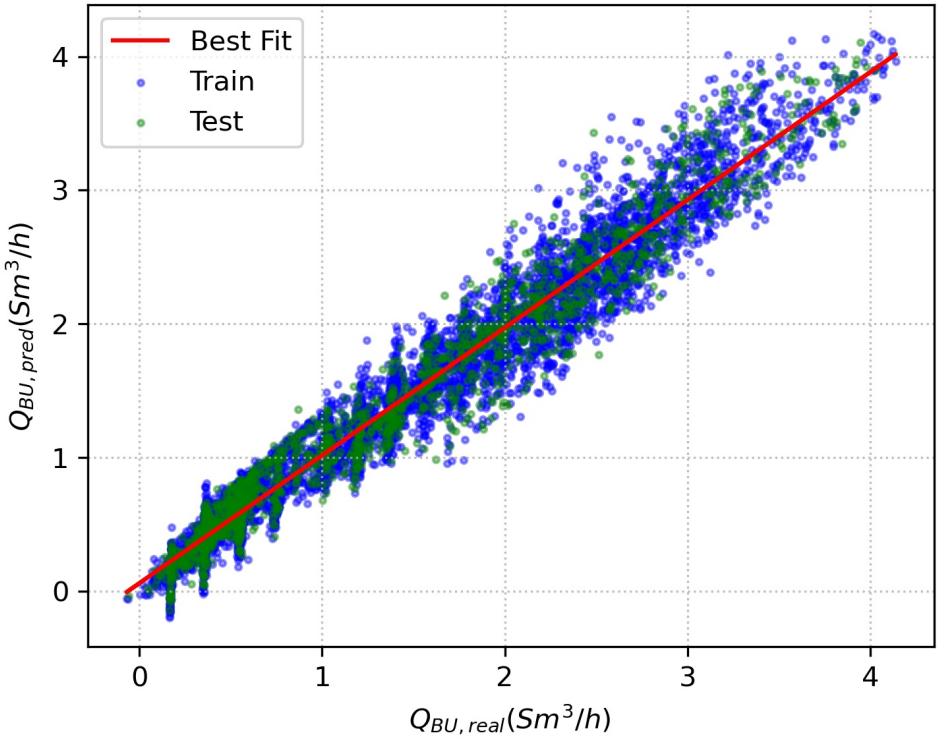


Figure 6.10: CGS_A MLR scatter plot: predicted vs real.

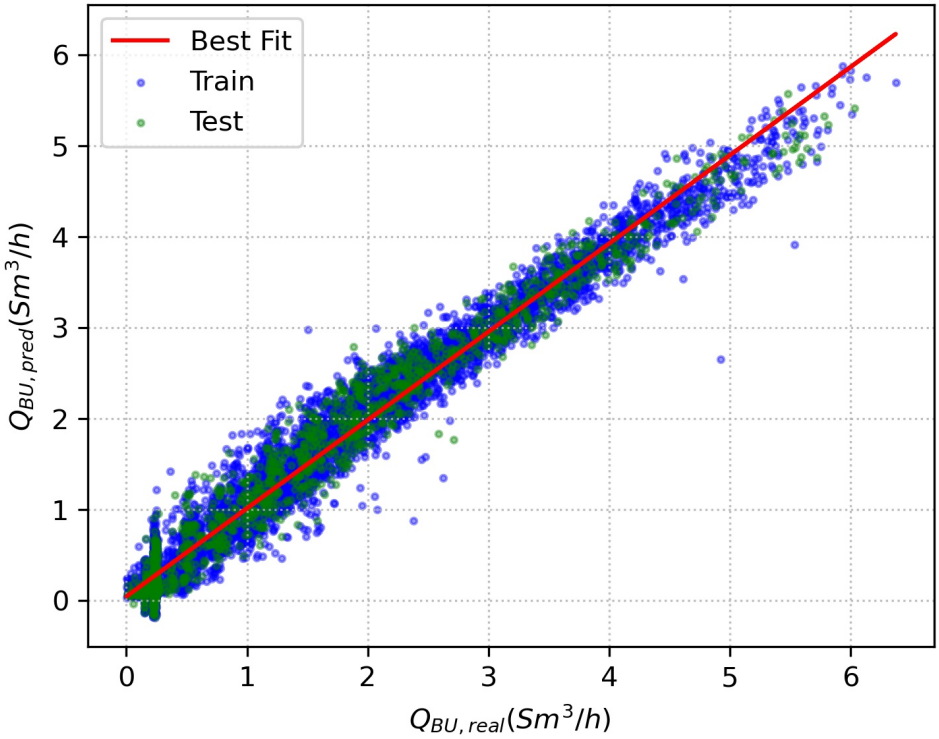


Figure 6.11: CGS_B MLR scatter plot: predicted vs real.

6.3 ANN models deployment

Both neural network models are trained and realized by relying on the sckit-learn libraries (cite scikit learn), in particular the MLPRegressor module. Using a full year, i.e. 365 days, the dataset is split as follows: 70% for training, 10% for cross-validation, and 20% for model testing evaluation. In this way,

80 % of the dataset is divided into 70% for actual training and 10% to go for model validation via the procedure integrated into the MLPRegressor function. The score of the validation dataset is used as a metric to block the training of the model before the over-fitting condition is realized. To train the models the MLPregressor early stopping is set to on. This is a form of regularization used to avoid overfitting when training a learner with an iterative method, such as gradient descent. Such methods update the learner to make it better fit the training data with each iteration. Up to a point, this improves the learner's performance on data outside of the training set. Past that point, however, improving the learner's fit to the training data comes at the expense of increased generalization error. Early stopping rules guide as to how many iterations can be run before the learner begins to over-fit.

Fine tuning with grid search

A model hyperparameter is a characteristic of a model that is external to the model and whose value cannot be estimated from data. The value of the hyperparameter has to be set before the learning process begins. For example, the number of hidden layers or hidden neurons for each layer in Neural Networks. Hyperparameters are set before training (before optimizing the weights and bias). Grid search is a process that searches exhaustively through a manually specified subset of the hyperparameter space of the targeted algorithm. Random search, on the other hand, selects a value for each hyperparameter independently using a probability distribution. Both approaches evaluate the cost function based on the generated hyperparameter sets. Among the hyper-parameters that were chosen for the realization of the neural network for each cabin, we have the following results:

- Activation functions: logistic, relu, tanh.
- Batch size: 5%, 10% and 20% of the training dataset number of data
- Number of hidden neurons: from 2 hidden neurons up to 300 hidden neurons.
- Learning rate

An exhaustive search over specified parameter values for an estimator has been chosen exploiting the library module of scikit-learn called GridSearchCV. To use that module you have to set the scoring metrics and the number of folds for the k-fold cross-validation. Cross-validation is a statistical method used to estimate the skill of machine learning models. It is commonly used in applied machine learning to compare and select a model for a given predictive modeling problem because it is easy to understand, easy to implement, and results in skill estimates that generally have a lower bias than other methods. For this work, we decided to use MAE for scoring and k =3 as number of folds for the k-fold CV. Here are the results of the grid search for both plants:

- CGS_A

  Activation functions: logistic

  Batch size: 10% of the training dataset number of data

  Number of hidden neurons: 10

  Learning rate: 0.005

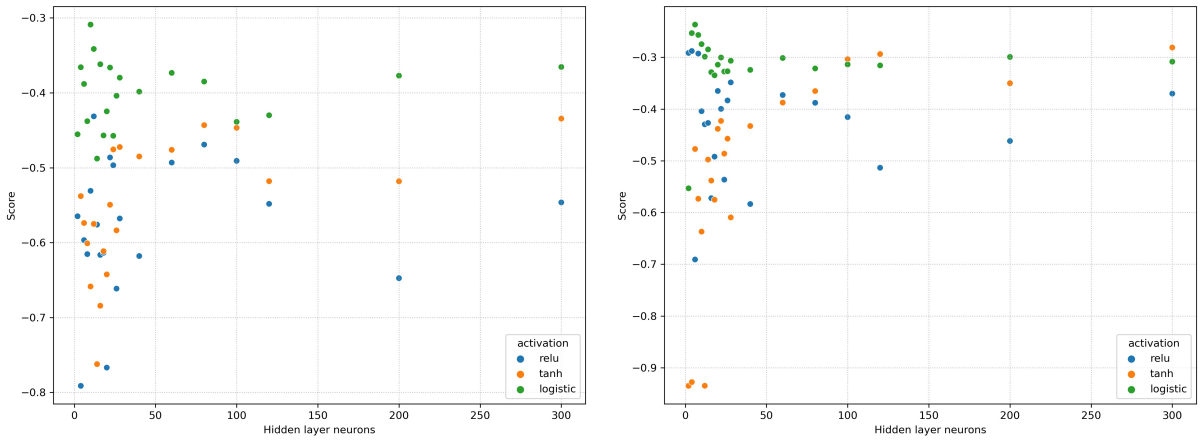
- CGS_B

  Activation functions: logistic

  Batch size: 5% of the training dataset number of data

Number of hidden neurons: 6

Learning rate: 0.0001



(a) (b)

Figure 6.12: Grid search results for CGS_A (a) and CGS_B: score of the model for three different activation functions and as the number of nuerons in the hidden layer varies

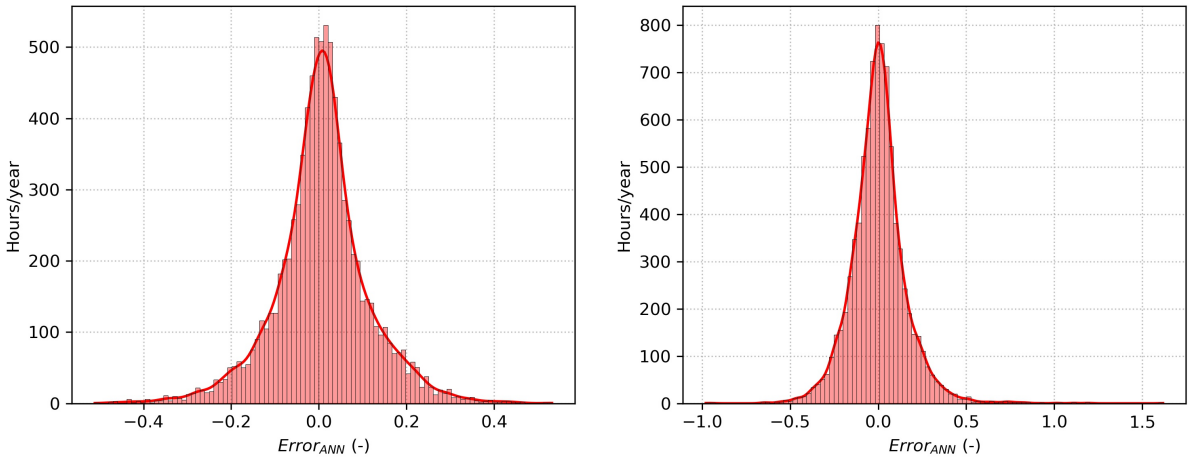
.

Figure 6.12 shows the mean scoring values of the models for plant A and B while searching at the same time the best activation function and the best number of hidden neurons of the one hidden layer. Both figures highlight the fact that the logistic function and the ReLu are performing better than the tanh activation function. It can bee seen that for small numbers of hidden neurons the logistic outperforms the ReLu, but their scoring is almost the same. While the number of hidden neurons increses, the logstic start to become better more and more.

Final models evaluation

This section presents the performance results of the neural network models for the two plants considered. The first table 6.5 shows the values of metrics related to the ANN training and testing dataset for the evaluation of CGS_A preheating consumption. Similarly to the case of MLR model traning, the dataset was divided into 80% training and 20% testing, but the training was divided into 70% actual training and 10% cross-validation. It is critical that in the final training the metrics are similar for both training and testing, which means that the model is performing and generalized and over or under fitting issues have been avoided. Figures 6.13a and 6.14 show the error distribution of the model training and the scatter plot of the predicted values vs the real values for the plant CGS_1. Blue shows the points related to the training set and green shows the points related to the test set. The ANN model of the first plant ultimately presented the following values of the metrics for the test set: MAPE of about 8.6%, MAE of 0.07, MSE of less than 0.012 and a RMSE of about 0.1 Smc/h. Moreover, the coefficient of determination is very high and close to 1 with coefficient values of 0.98. Figures 6.13b and 6.14 show the error distribution of the ANN model training and the scatter plot of the predicted values vs the real values for the plant CGS_B. The ANN model of the second plant ultimately presented the following values of the metrics for the test set: MAPE of about 21%, MAE of 0.11, MSE of less than 0.02 and a RMSE of about 0.15 Smc/h. Moreover, the coefficient of determination is very high and close to 1 with coefficient values of 0.98. Both ANN models outperformed prediction performance on the test set compared with their respective MLR rivals for both plants. This confirms what was expected from the beginning of the work: the ANN model allows to capture the inherent nonlinearity of these types of implants, whereas a simple MLR model fails

to do so. Very important to note that the distribution of ANN model errors with respect to the overall training dataset is extraordinarily more symmetrical and balanced than in the MLR case. It means that in those cases the ANN models managed to be predictive and robust without going over-estimating or under-estimating.



(a) (b)

Figure 6.13: (a) CGS_A ANN errors distribution vs (b) CGS_B ANN errors distribution

Table 6.5: CGS_A : ANN final model metrics

	Unit	Test	Train	Train + Test
MAPE	[%]	8.066373	8.71887	8.588371
MSE	$[Smc/h^2]$	0.012347	0.01134	0.011541
MAE	[-]	0.078659	0.076668	0.077067
RMSE	[Smc/h]	0.111116	0.106489	0.10743
RMSE%	[%]	2.759271	2.573446	2.633263
R^2	[-]	0.986961	0.987549	0.987432
Size	[%]	20.00%	80.00%	100.00%

Table 6.6: CGS_B : ANN final model metrics

	Unit	Test	Train	Train + Test
MAPE	[%]	21.04828	22.9781	22.592094
MSE	$[Smc/h^2]$	0.023792	0.02477	0.024574
MAE	[-]	0.110912	0.111634	0.11149
RMSE	[Smc/h]	0.154247	0.157385	0.156762
RMSE%	[%]	2.556297	2.469171	2.657338
R^2	[-]	0.987327	0.986597	0.986746
Size	[%]	20.00%	80.00%	100.00%

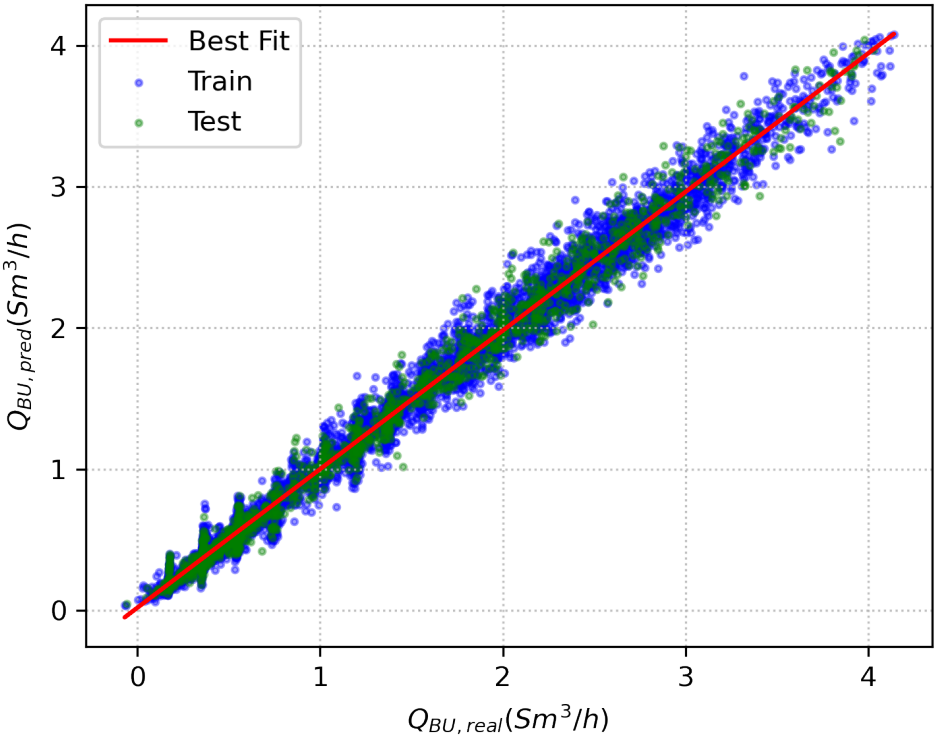


Figure 6.14: CGS_A ANN scatter plot: predicted vs real.

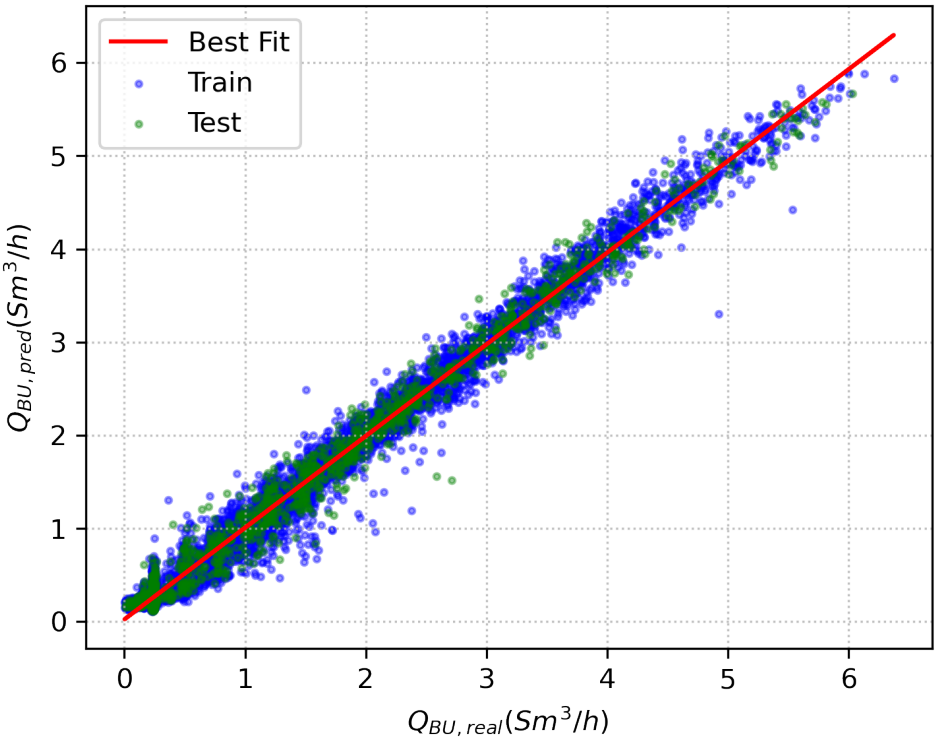


Figure 6.15: CGS_B ANN scatter plot: predicted vs real.

6.3.1 ANN models prediction comparison with real data

The DSOs, and in particular the partner who provided the data used, do not currently use predictive models to predict preheating energy consumption. Generally, what is done by the DSO is to go to the plant site and periodically monitor for malfunctions or energy wasting events.It is an on-call, non-condition-based

type of maintenance of the thermal power plant operation. To implement condition-based maintenance, it is necessary to implement models to understand the "state" of the health of the plant from the perspective of energy performance. Considering the extremely promising results of the developed models, they can be implemented directly on the machine, and their value is compared with the real value read by the system. An example of the results is shown in the following figures 6.16 and 6.17. The two figures show the trends of the ANN models for the two plants compared with the real values for the 4 seasons relating to the training period. There is an excellent correspondence both in terms of trend and in terms of absolute value for both systems and the related models developed in this chapter. For both systems, it can be clearly seen that the two graphs for the most wintry season, i.e., April 15 and February 15, is perfectly replicated in terms of model predictions, compared to the actual data. A first direct application of the newly developed models could be to use them directly in comparison with real data, to implement an instantaneous "anomaly detection" system. Figure 6.18 and figure 6.19 the results of this implementation for 4 dates after the start of the testing period. The confidence bands of the predicted model are plotted, to be able to experiment with a first use of the algorithm compared to the real case. The width of the tolerance interval is obtained by multiplying the standard deviation of the training error of the ANN models of each plant by a safety coefficient, which for this work was set equal to 3 The following (6.4) is used to compute the above-mentioned confidence interval.

$$\epsilon_{confidence} = \sigma_{ANN,errors} \times SC \tag{6.4}$$

This value may be tuned by the DSO to include more or fewer alarms within its alarms history log. It is noted that, although the models are extremely performing, for a few days alarms would have arisen due to the curve of real values which completely goes outside the tolerance bands shown in the figure. For both plants, problems would be encountered in both May and November and the DSO could be alerted to the discrepancy between the model and real data. However, in both cases, no malfunction was found by the DSO. This is because this type of approach is best suited to plants that risk catastrophic breakdowns or failures that are very frequent and which can be analyzed by instantly comparing a parameter only with its digital twin counterpart. In addition, because there is a strong dependence on seasonality , each day may be very different from the previous one or the same one a year earlier. To better use the models developed based on a "baseline" period, it would be more appropriate to evaluate the development of errors over time and not just the value of the error over the course of a single day. Therefore, please refer to the development of the CuSUM-based energy monitoring method in the next chapter. This would allow an analysis of what the DSO is actually interested in:

- to perceive whether the system is drifting toward a malfunctioning and thus energy-wasting condition
- knowing the date of system upgrades or refurbishments, whether the energy improvement intervention is bringing benefits over time.

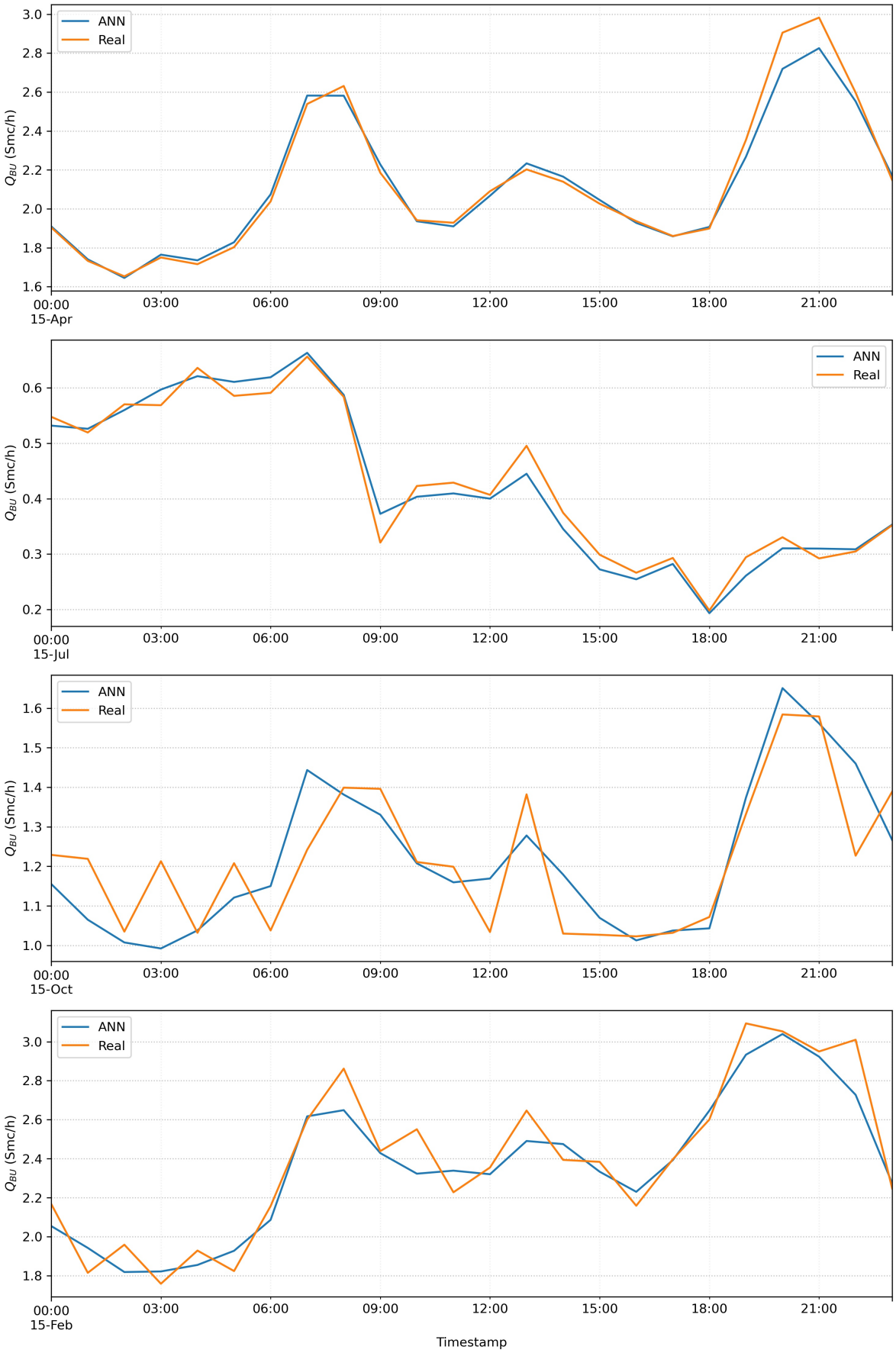


Figure 6.16: CGS_A: comparison results between ANN and real data for training period

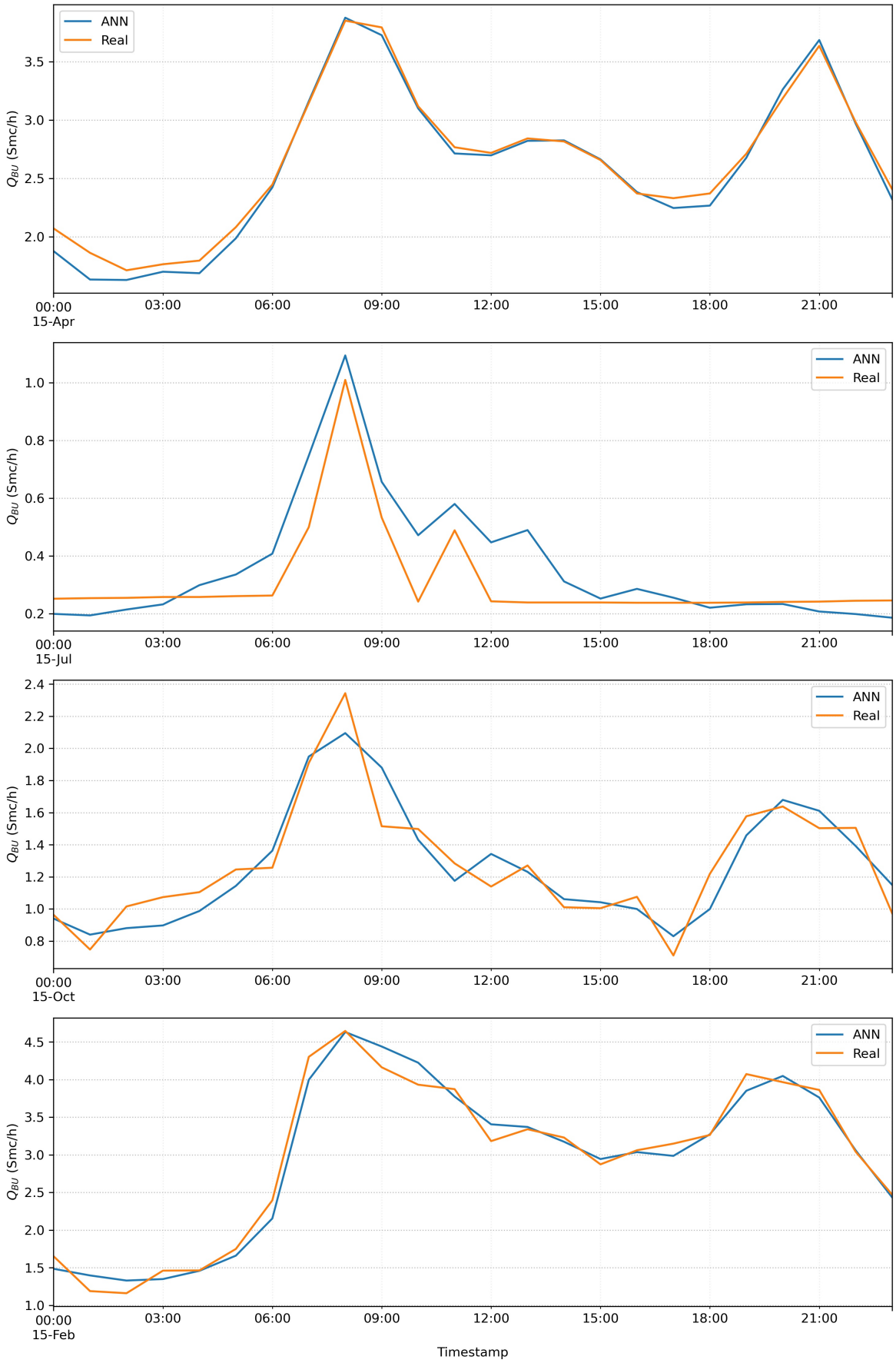


Figure 6.17: CGS_B: comparison results between ANN and real data for training period

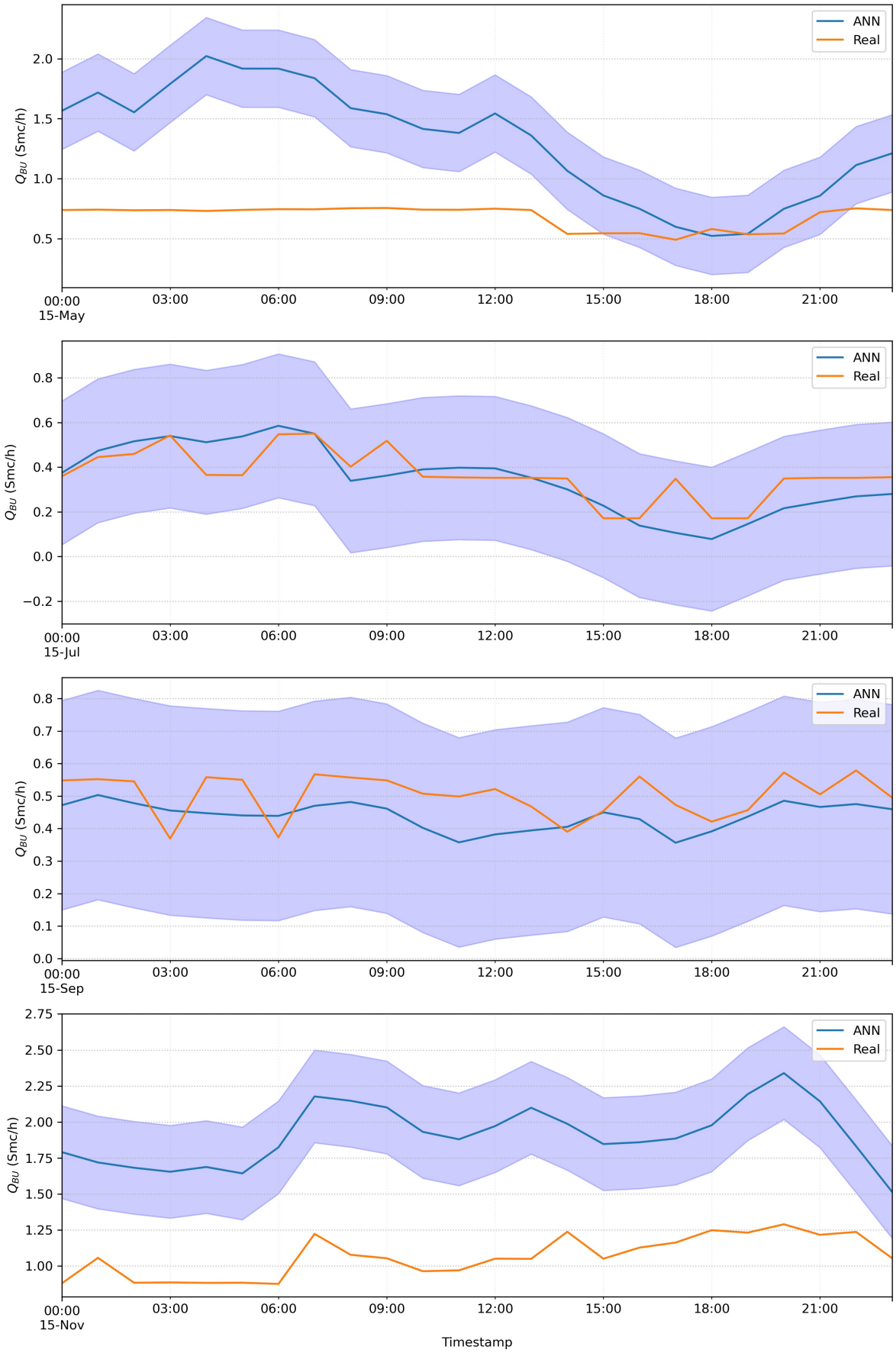


Figure 6.18: CGS_A: comparison results between ANN and real data for testing period with confidence intervals

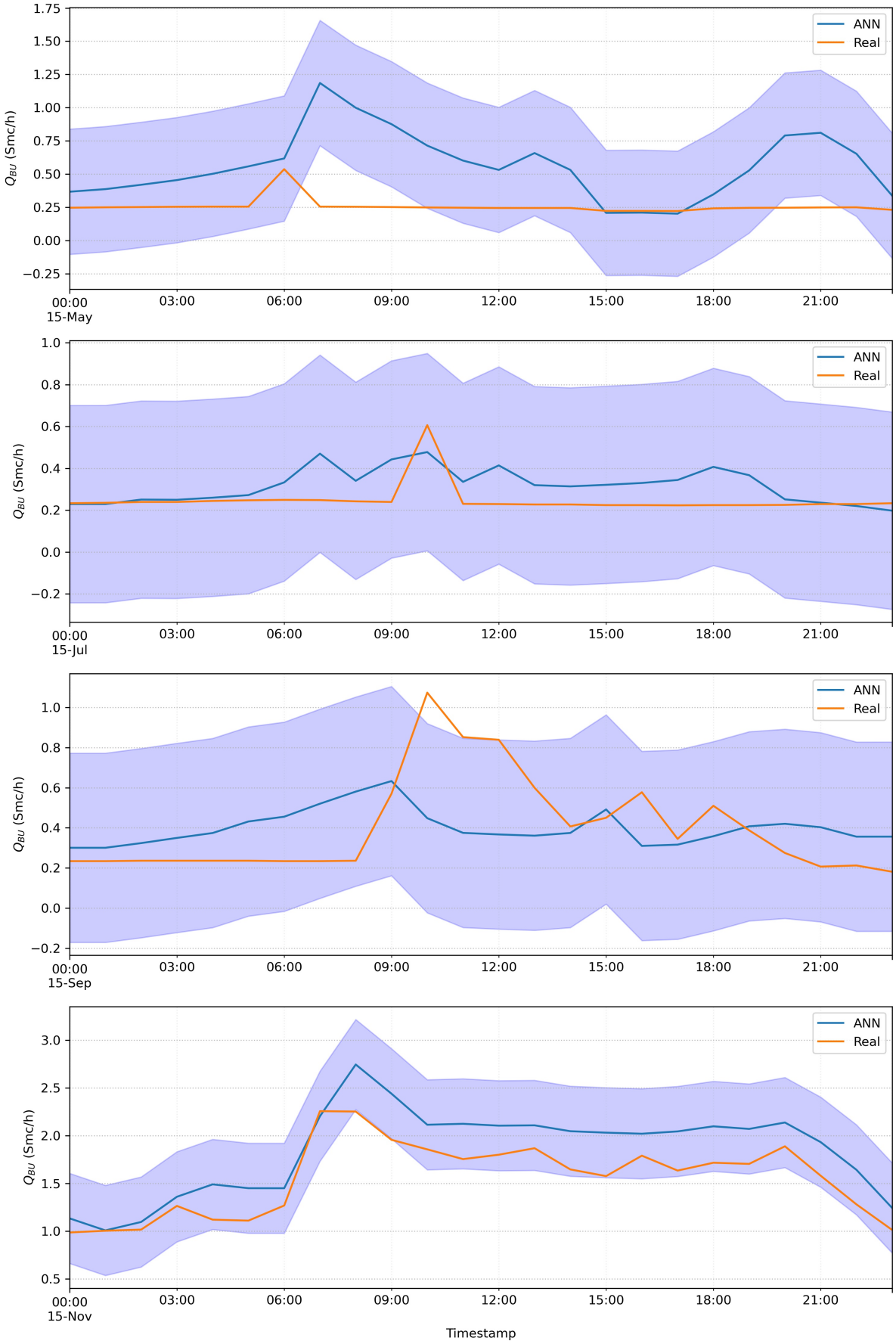


Figure 6.19: CGS_B : comparison results between ANN and real data for testing period with confidence intervals

6.4 Energy monitoring results

This paragraph shows the trend of the CuSUM graphs obtained with the regression models developed for this work for both plants. All CuSUM curves are calculated for the entire duration of the dataset, from the first day of training (15 April 2021) to the last day of training (15 April 2022), until the end of the testing period at disposal (15 December 2022). It is immediately noticeable how all four CuSUM curves, whether obtained as the difference between real and MLR models or real and ANN models, return towards zero at the end of the training period (highlighted by a dashed green line); that is, it is a testament to the goodness of the models developed that by definition they reduce the quadratic error between real data and output throughout the training dataset. Nevertheless, as was to be expected after analyzing the distribution of training errors for both models, the MLR models are clearly more unbalanced than the ANN models. They present approximately double the standard deviation of errors for both systems; moreover, for the former figure 6.20 there is an effective imbalance for almost the entire duration of the training which is only reduced at the end of the aforesaid period. Figure 6.21 shows the comparison of the two cusum obtained with the MLR and the ANN trained with the CGS_B datasets.

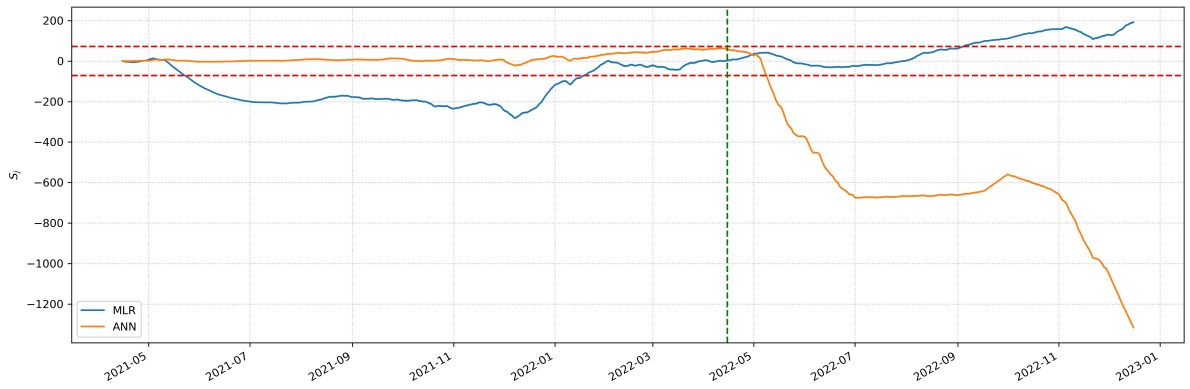


Figure 6.20: CGS_A CuSum results: MLR vs ANN

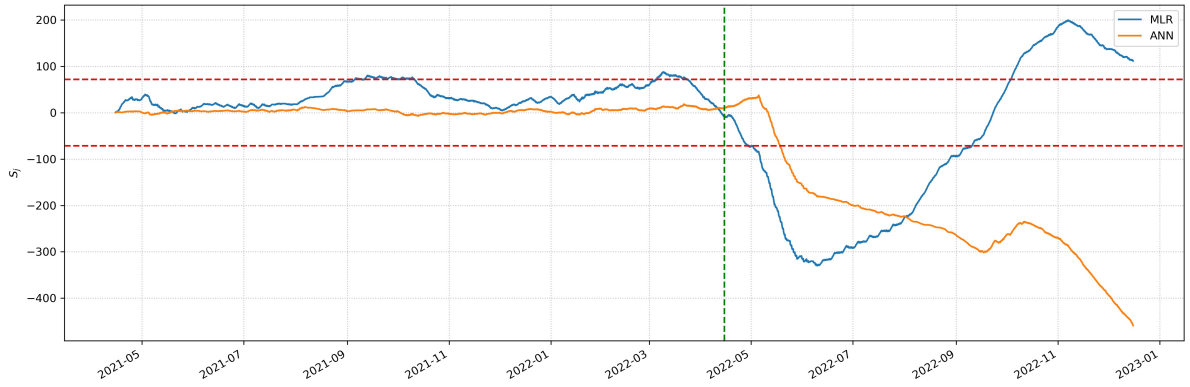


Figure 6.21: CGS_B CuSum results: MLR vs ANN

The hypothesis underlying energy monitoring using CuSum techniques is the reliability and robustness of the "baseline" model; in this case it is clear that the neural network is clearly more performing than the MLR and therefore it will be used as a test case by comparing the trends of the ANN curves with respect to what actually happened in the plant. To better explain this misalignment between the MLR CuSum and the ANN CuSum, it could be useful to plot the errors during the testing phase, then as of mid-April 2022 in Figures 6.22 and 6.23. For plant A, the greater robustness and accuracy of ANN's model is evident from the average of errors for a long period equal to zero. Where the heat pump has

been turned on, however, i.e., the first few months of testing, both models notice this and in fact the error is stably negative, meaning that the CuSuM is decreasing and the model predicts higher consumption than what actually happened. After this, however, the two models diverge, in that the MLR shows a trend of stably positive errors, thus indicating potential extra consumption, while in contrast, the ANN is stably on zero mean errors, confirming that nothing significant has happened in the plant. While then from November onward this trend of misalignment re-couples and in fact, both cusums go back down as do the errors.

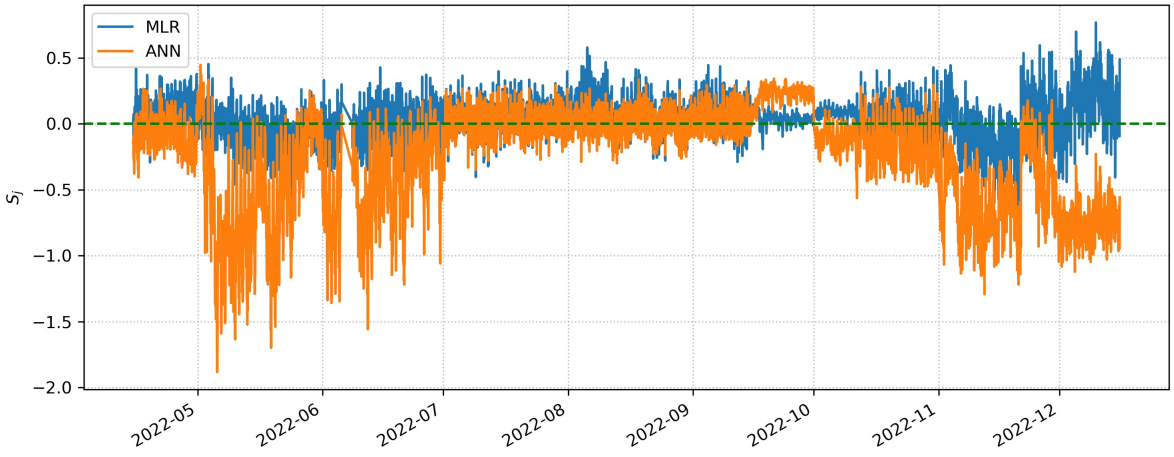


Figure 6.22: CGS_A: Errors MLR vs ANN.

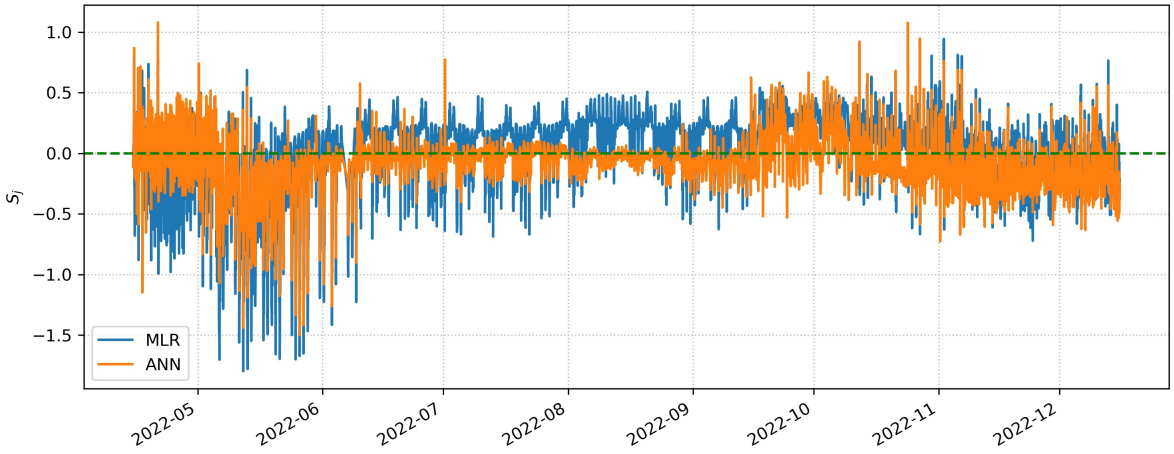


Figure 6.23: CGS_B: Errors MLR vs ANN.

The two following figures, 6.24 and 6.25, show the trend of the CUSUM given by the ANN models only for both plants.

This is what can be deduced from the analysis of the CUSUM trend 6.24 for the first plant.

- May - July: the operators carried out a recalibration of the pre-heating water temperature (see figure 4.11 in chapter 4) and this meant that the predicted consumption was higher than the actual one. The model has no information in the inputs relating to the water set point or its value and consequently has predicted a higher consumption than expected.
- July - October: no variation detected with CuSum method.
- October - end of Dataset: another drop in water temperature was detected which led to a reduction

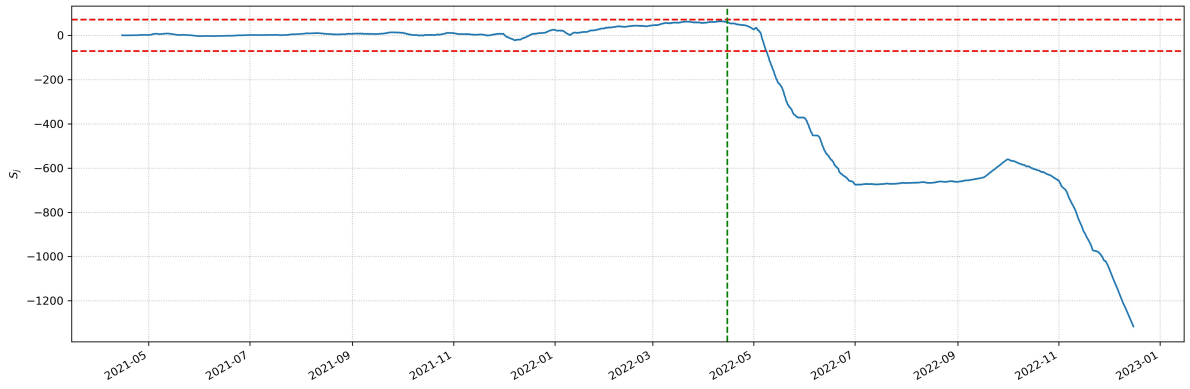


Figure 6.24: CGS_A Cusum results with ANN.

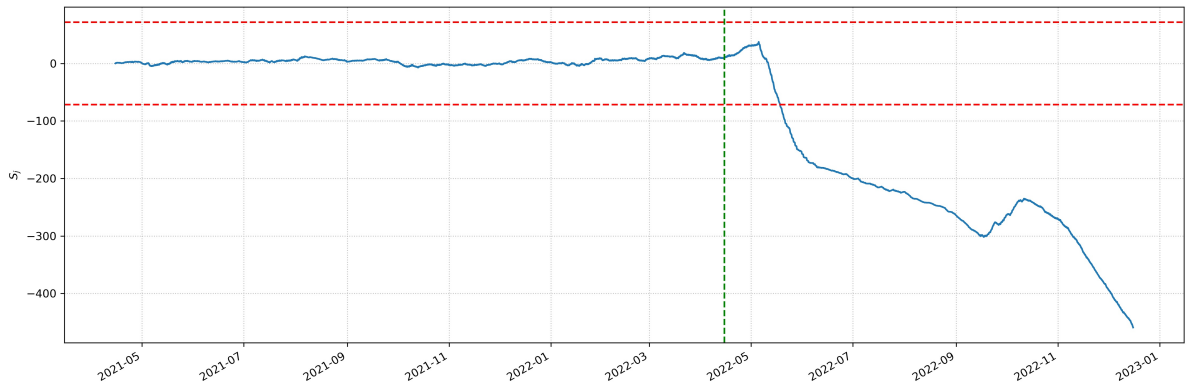


Figure 6.25: CGS_B Cusum results with ANN.

in gas consumption, probably this time due to a possible malfunction of the thermal plant (see figure 4.11 of chapter 4).

Cross-referencing the cusum trend of Figure 6.25 and comparing with the DSO in order to assess the effects of plant interventions or to ascertain plant malfunction shows that:

- Early May - early June: the CuSum becomes negative with a very marked slope, this would agree with the effect of turning on the heat pump relative to that period leading to a reduction in gas consumption under the same conditions and thus a negative CuSum, because Y_{real}- $Y_{pred} < 0$.
- early June - late November : the CuSum slows its slope and has both positive and negative fluctuations around 3 times the standard deviation ell error of the training period. However, in this period (which is the summer period) there was a change in the natural gas set point temperature, which is usually kept at 10°C in the summer period, but for some reason it was kept lower and specifically at 8°C. The model, trained with a perfect set point scheduling in accordance with what is also described in the previous chapters, would expect a higher consumption with a lower set point and instead the consumption is lower and consequently the CuSum continues to decrease albeit in a less marked way than in the previous phase.
- late October - end of the dataset: without subsequent data available, it is not possible to extract much information, however the model, with its CuSuM variation has noticed a sudden change in water temperature compared to the previous year. At the end of September the water temperature is higher than the previous year while at the end of October the water temperature is about 10-15°C lower than the previous year; this is perceived by the model which signals with the cusum a negative

drift in consumption due to a lower water temperature. Problem with the water temperature set point.

6.5 Concluding remarks

Several regression-based machine learning models were developed and trained, a Multiple Linear Regression (MLR) and an Artificial Neural Network (ANN) algorithms. The selection of the inputs for these models is obtained through the features selection and engineering process. The algorithms were trained on a training data set relating to a complete year of operation for two real plants and their predictive performances were tested on another data set relating to the subsequent operating conditions of the same plants. The algorithms performed greatly, in terms of all the metrics chosen for their evaluation on the testing dataset. The algorithms, and in particular the ANN models, which are capable of understanding the non-linear interactions of the system, performed extremely satisfactorily on both datasets in terms of all the metrics chosen for their evaluation, such as R^2, RMSE. Values higher than 0.98 were obtained for the R^2 coefficient, with mean square percentage errors lower than 3% for both models and plants. The models were integrated into an energy monitoring system using the CUSUM technique and it was possible through these to identify malfunctions and waste within the testing period. The main outcomes of this first part of the work are listed below:

- The filter features selection method meant discards T_{wat} , T_{gas}, $Daytype$, $Weekday$ and the following variables were kept: Q_{gas} , P_{in} , P_{out} , $T_{SP,wat}$, T_{SP}, T_{env} , G, $Month$, $Sin(H)$, $Cos(H)$, $T_{gas,in}$. The features engineering process is fundamental for this type of application. The addition of temporal and seasonal variables, such as the month of the year or the time of day in sinusoidal form, make it possible to improve the performance of the models by up to 20% in the case of MLRs for the second cabin, considering the RMSE score.

- Best ANN models reach values of about 0.98 for the R^2„ 0.11 and 0.15 Smc/h for the RMSE and 2.5-2.7 % errors for the percentage RMSE for the testing dataset for both plants.

- Neural network models perform better both in terms of absolute and percentage errors on both the training and reference test sets, compared to MLR models. Therefore, they are chosen as templates to create the energy “baseline” curve to be used in the cusum-based method for energy monitoring.

- The process of direct implementation of the models as energy anomaly detection systems is not sufficiently robust to capture consumption drifts or for any assessments of the impact of efficiency actions. The CuSUM method, on the other hand, allows for the creation of "baseline" consumption models and to evaluate of the performance of the CGS over time to find the variations between the actual and the modeled gas consumption, being an essential tool for monitoring the effectiveness of the natural gas preheating system.

Chapter 7

Conclusions

This thesis work consisted of two successive steps, the common theme of which was the energy analysis and the possibility of efficient natural gas pre-heating systems within the City Gate Stations in the gas distribution network in the Italian Scenario. A third element, preparatory and in support of the previous two just mentioned, was the analysis of real datasets provided by the DSOs, in particular 7 datasets of real plants ranging in size from very small to medium-large. In addition, for two of these plants, in particular the first two, datasets on gas preheating systems were also provided. The analysis of the datasets showed first of all that working with these systems means having to find correlations and commonalities between variables within a plant fleet that is extremely varied in terms of plant layouts and operating conditions. While it is possible to have a variety of configurations concerning the flow rates processed and the pressure drops to be managed in terms of their annual distributions, it was found that the dependence of the gas flow rate on the outside temperature follows a similar trend for all plants; this is because the gas consumption of the distribution network analyzed is strongly influenced by seasonality and consequently by the outside temperature. Therefore, it can be seen that CGS analysis is a problem that is strongly influenced by seasonality and plant operating conditions in terms of processed flow profiles, gas pressure drop values, and plant layout configuration. First of all, the estimation of preheating using lumped parameter models or models based on first principles does not allow for satisfactory results in absolute terms, but very good ones in terms of quality and consumption profile. This is unfortunately due to the almost impossibility of generalizing simple equations and applying them to a CGS park that includes many different configurations with even different technologies and operating conditions that are never the same; moreover, the estimation of the losses would require a validation effort not repaid in any case to analyze decarbonization scenarios. Among the various technologies available to implement the energy efficiency and decarbonization process of preheating in CGS, it was decided to focus on air source heat pumps for several reasons: high efficiency, low capital costs when compared to expansion or other types of solutions, the possibility of rapid installation and immediate decarbonization of gas and, finally, possibility of access to energy efficiency certificates by the DSO. Starting from the assumption that without a seasonal storage solution and assuming that only the best techno-economic solutions are considered, which are feasible by the DSOs in practice, it was not possible to obtain a 100% decarbonization for all the systems analyzed as case studies. However, the results show that it is possible to obtain a return on investment in a short time, from 6 to 8 years or around 4-5 years respectively for the control logic which does not envisage powering the heat pumps even with electricity purchased from the network and for the one that provides for the purchase of electricity from the national network. The sensitivity analysis showed that the techno-economic feasibility of these solutions is very susceptible to gas prices in absolute terms and, in a much less impactful way, to the variation of a gas and electricity price ratio. Assuming that the

resale price of electricity, known as the minimum withdrawal price, is constant over time and low and in such a way as not to transform the purpose of the solution from a gas consumption reduction and efficiency system to an electricity production system, the first two plant logics (no purchase from the network) are extremely penalized if the price of gas settles on pre-crisis values and lower than around 0.6-0.5 €/Smc. At the same time, the solution that provides for the possibility of purchasing from the network guarantees a much greater resilience of the solution and its techno-economic feasibility as gas prices vary. This is mainly due to two factors: the reduction in gas prices is somehow balanced by the fact that a significant share of gas is still being saved by purchasing electricity at a lower cost and producing a quantity of heat three times, on average, compared to the electricity purchased and why, the ratio between gas price and electricity price would also increase, favoring this solution even more. Finally, the results of the first analysis show that the maximum decarbonization percentage obtainable is on average less than 30% of the annual preheating demand, if an off-grid or partially on-grid control logic is used, or higher decarbonization can be obtained even if partially green, when the electricity grid is not yet 100% zero impact, even up to 80% of the total energy of the preheating request if a bidirectional on-grid logic is used. The second analysis, i.e. the feasibility study of implementing machine learning algorithms for energy monitoring purposes, was tested on two case studies, which envisaged the presence of datasets of around two years relating to the pre-heating systems of two small-scale sizes provided by the DSO. The correlation analysis of the two datasets showed that, as regards the raw features, i.e. the sensors installed in the system, the most marked correlation with the target is given by the gas flow rate passing through the CGS, the external temperature, and any control of the preheating plant, if present. This certifies the strong dependence of the target on seasonality, which in turn is due to the composition of users downstream of the plant. Several variables also have a greater Spearman coefficient value than the Pearson coefficient value: this means that the dependence between them and the target is monotonic, but also non-linear, such as between the ambient temperature and the gas flow rate and therefore consumption. This is very important as it affects the performance of the regression models used in the next step. The implementation of the Multiple Linear Regression models has made it possible to obtain already satisfactory results in terms of regression metrics and to highlight the positive effect of the inclusion in the dataset of engineered features, i.e. those obtained by extrapolating new information from the raw data with the help of the physical knowledge of the system. The final MLR model therefore performed better than the baseline MLR model, increasing the coefficient of determination R^2 by an average of 2% and reducing the MAE by 20% and by 40%, the RMSE by 11% and by 22 % for the two plants, respectively. The gas inlet temperature and the month of the year for both plants have been included among the candidate features, as proof of the importance of the gas inlet conditions the seasonality of consumption, and the time of day of the sinusoidal shapes for the second cabin. The inclusion of a neural network model, which unlike the MLR allows grasping the non-linear relationships of a regression problem between targets and features, has paved the way for significant improvements in the performance of the model and made the error distribution much more homogeneous during training and consequently have a more generalizable model. This is also reflected in the final energy monitoring analysis: the ANN is more stable during the CUSUM count in the training phase, with lower standard deviations of the error, and allows to analyze of the behavior of the plants in a more robust way in the testing phase. The cusum analysis showed that by using complex, non-linear machine learning algorithms, such as FF-MLP neural networks, it was possible to obtain a reliable consumption baseline during the training period. During the testing period, changes were highlighted in the real systems analyzed, such as the switching on of a heat pump for efficient pre-heating of the system or the adjustment of the hot water and gas set point temperatures. In conclusion, this thesis work has made it possible to answer the research questions that had initially been posed: simplified

modeling of the CGS and its preheating requests based on physical models allows us to have useful tools in the analysis of decarbonization scenarios and therefore to be a good planning tool for all DSOs who want to maintain their decarbonization commitments and possibly access tax relief related to the purchase and sale of energy efficiency certificates. However, the presence of a very varied park in terms of configurations and clusters of users downstream and the marked dependence on the summer/winter season makes these plants difficult to completely decarbonize through systems that do not provide for seasonal storage and based on production from renewable energies such as the photovoltaic. Secondly, the analysis of these plants from the point of view of real datasets makes it possible to understand which are the parameters most influencing the systems and the correlation between consumption and the rest of the variables. The data-driven modeling problem collided with the inherent non-linearity of the physical system that was analyzed but allowed us to understand that with a sufficiently predictive and generalizable regression model, it is possible to monitor the energy operating status of the systems and possibly also the improvements. The impact of the work is significant and could be summarised as follows: the work has led to the creation of a rapid approach that can be easily integrated and scaled by any DSO to enable rapid decarbonization of the subsystem under consideration. Thermodynamic models are easily generalizable and scalable and do not require a great deal of computational effort, but they are sufficiently accurate tools for techno-economic feasibility analysis to immediately propose solutions to reduce and contain consumption, avoiding focusing on particular plants that are difficult to implement in practice. Data-driven analysis makes it possible to propose methods to go about creating energy baseline models very quickly, as all DSOs have access to the histories of all the plants they control and could put in place a facilitated plant modeling campaign based on the results of the study already done, overcoming current manual or rule-based procedures. All this allows to support and help DSOs, providing them with a series of tools that can be useful to address, acting immediately, the challenge of decarbonizing gas infrastructures.

Appendix A

Published papers

A.1 Offline Monitoring Method for a Natural Gas City Gate Station Odorization System

Year: 2022 SITE: https://onepetro.org/PSIGAM/proceedings-abstract/PSIG22/All-PSIG22/485454

A.1.1 Abstract

Natural Gas (NG) is odorless and therefore requires an odorizer to be injected into its flow to ensure detection when a gas leak occurs and thus provide satisfactory safety levels. The odorisation process is a delicate step that takes place within the City Gate Stations (CGS), which are key elements of the NG network infrastructure. This work aims to develop a method for offline monitoring of the odorization process within a CGS located in central Italy, based on the exploiting of the odorization station dataset through several machine learning models development, to evaluate the odorization process performance. An unsupervised machine learning method based on two different algorithms, the LOF (Local Outlier Factor) algorithm and the K-Means clustering, is developed, and then data mining is carried out on the dataset to extract useful information. The results show that the use of the algorithm made it possible to identify anomalous points in the dataset and their dependence on the main operating parameters of the CGS, as well as some clusters of under-odor and over-odor tendencies for the system under consideration.

A.2 Data-driven modelling for gas consumption prediction at City Gate Stations

Year: 2022 DOI: 10.1088/1742-6596/2385/1/012099

A.2.1 Abstract

City Gate Stations (CGS) are critical elements of Natural Gas (NG) distribution systems, as they connect national high-pressure transmission networks with local low-pressure networks. One of their main tasks is to pre-heat the gas to avoid dangerous sub-cooling due to the Joule-Thompson effect after the pressure reduction stage. For this process, significant amounts of thermal energy are required, usually obtained by burning part of the gas flow rate. This work aims to develop a data-driven model that will serve as a tool to predict and monitor the thermal consumption of the CGS. The plant chosen as a case study for this activity is in a region of central Italy. A Multiple Linear Regression (MLR) model is developed

and trained, and its predictive performance is evaluated. The model results achieved an accuracy of over 95% for the coefficient of determination. The method makes it possible to create a baseline consumption model and evaluate the performance of the CGS over time using the CUSUM technique to find variations between actual and modelled gas consumption, being an essential tool for monitoring the effectiveness of the NG preheating system.

A.3 Integration of Renewable Energy Systems at City Gate Stations to Reduce Pre-Heating Gas Consumption

Year: 2023 DOI: https://doi.org/10.13044/j.sdewes.d11.0447

A.3.1 Abstract

The Italian Natural Gas (NG) distribution network includes thousands of NG metering and pressure reduction stations, called City Gate Stations (CGS), for injecting gas into low-pressure networks. These plants are mainly based on the isenthalpic throttling of the gas flow to reduce its pressure, which leads to a significant reduction of its temperature by the Joule-Thompson effect. Gas preheating systems that avoid excessive cooling are installed upstream of pressure reduction valves and usually exploit conventional gas boilers. The energy consumption and carbon footprint could be reduced by integrating heat pumps coupled with renewable energy sources for NG preheating to support gas boilers. For this work, an ad-hoc thermodynamic model for estimating the thermal energy demand for pre-heating is developed, exploiting experimental data from a real CGS, and simplified models of heat pumps and renewable systems. This work aims to assess the actual technical and economic feasibility of energy savings through these technologies. Results show the validated model to be sufficiently accurate to estimate the need for gas preheating for these applications. For the considered case study, up to 38%, 32%, or 26% of the total thermal energy can be recovered with a PBT of less than 20 years, 15, and about 13 years, respectively.

A.4 Other papers

- L. Cheli and C. Carcasci, 'Modelling and analysis of a liquid-cooled system for thermal management application of an electronic equipment', E3S Web Conf., vol. 197, p. 10008, 2020, doi: 10.1051/e3sconf/202019710008.
- C. Carcasci, L. Cheli, P. Lubello, and L. Winchler, 'Off-Design Performances of an Organic Rankine Cycle for Waste Heat Recovery from Gas Turbines', Energies, vol. 13, no. 5, Art. no. 5, Jan. 2020, doi: 10.3390/en13051105.
- L. Ciappi, L. Cheli, I. Simonetti, A. Bianchini, G. Manfrida, and L. Cappietti, 'Wave-to-Wire Model of an Oscillating-Water-Column Wave Energy Converter and Its Application to Mediterranean Energy Hot-Spots', Energies, vol. 13, no. 21, Art. no. 21, Jan. 2020, doi: 10.3390/en13215582.
- L. Cheli, G. Guzzo, D. Adolfo, and C. Carcasci, 'Steady-state analysis of a natural gas distribution network with hydrogen injection to absorb excess renewable electricity', International Journal of Hydrogen Energy, vol. 46, no. 50, pp. 25562–25577, Jul. 2021, doi: 10.1016/j.ijhydene.2021.05.100.
- L. Cheli and C. Carcasci, 'Model-Based Development of a Diagnostic Algorithm for Central Inverter Thermal Management System Fault Detection and Isolation', in 2021 5th International

Conference on System Reliability and Safety (ICSRS), Nov. 2021, pp. 14–21. doi: 10.1109/ICSRS53853.2021.9660763.

- L. Ciappi et al., 'Wave-to-wire models of wells and impulse turbines for oscillating water column wave energy converters operating in the Mediterranean Sea', Energy, vol. 238, p. 121585, Jan. 2022, doi: 10.1016/j.energy.2021.121585.

- G. Guzzo, L. Cheli, and C. Carcasci, 'Hydrogen blending in the Italian scenario: Effects on a real distribution network considering natural gas origin', Journal of Cleaner Production, vol. 379, p. 134682, Dec. 2022, doi: 10.1016/j.jclepro.2022.134682.

Bibliography

[1] 'IEA – international energy agency', IEA. (), [Online]. Available: https://www.iea.org/data-and-statistics.

[2] 'Topic: Natural gas industry in italy', Statista. (), [Online]. Available: https://www.statista.com/topics/9209/natural-gas-industry-in-italy/.

[3] 'Snam home page'. (2022), [Online]. Available: https://www.snam.it/it/index.html.

[4] 'UNI 9165:2020 - UNI ente italiano di normazione'. (), [Online]. Available: https://store.uni.com/uni-9165-2020.

[5] 'ARERA - home page'. (), [Online]. Available: https://www.arera.it/it/index.htm.

[6] M. Cavana and P. Leone, 'Biogas blending into the gas grid of a small municipality for the decarbonization of the heating sector', *Biomass and Bioenergy*, vol. 127, p. 105 295, 1st Aug. 2019, ISSN: 0961-9534. DOI: 10.1016/j.biombioe.2019.105295.

[7] N. Keogh, D. Corr, R. O'Shea and R. F. D. Monaghan, 'The gas grid as a vector for regional decarbonisation - a techno economic case study for biomethane injection and natural gas heavy goods vehicles', *Applied Energy*, vol. 323, p. 119 590, 1st Oct. 2022, ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2022.119590.

[8] P. Colbertaldo, G. Guandalini and S. Campanari, 'Modelling the integrated power and transport energy system: The role of power-to-gas and hydrogen in long-term scenarios for italy', *Energy*, vol. 154, pp. 592–601, 1st Jul. 2018, ISSN: 0360-5442. DOI: 10.1016/j.energy.2018.04.089.

[9] L. Cheli, G. Guzzo, D. Adolfo and C. Carcasci, 'Steady-state analysis of a natural gas distribution network with hydrogen injection to absorb excess renewable electricity', *International Journal of Hydrogen Energy*, vol. 46, no. 50, pp. 25 562–25 577, 21st Jul. 2021, ISSN: 0360-3199. DOI: 10.1016/j.ijhydene.2021.05.100.

[10] G. Guzzo, L. Cheli and C. Carcasci, 'Hydrogen blending in the italian scenario: Effects on a real distribution network considering natural gas origin', *Journal of Cleaner Production*, vol. 379, p. 134 682, 15th Dec. 2022, ISSN: 0959-6526. DOI: 10.1016/j.jclepro.2022.134682.

[11] G. Guzzo, L. Busi and C. Carcasci, 'Intrinsic leakage reduction of a natural gas distribution network by regulating the outlet pressure of a city gate station', presented at the PSIG Annual Meeting, OnePetro, 10th May 2022.

[12] M. Farzaneh-Gord, A. Arabkoohsar, M. D. Dasht-bayaz and V. Farzaneh-Kord, 'Feasibility of accompanying uncontrolled linear heater with solar system in natural gas pressure drop stations', *Energy*, vol. 41, no. 1, pp. 420–428, May 2012, ISSN: 0360-5442. DOI: 10.1016/j.energy.2012.02.058.

[13] M. Farzaneh-Gord, R. Ghezelbash, M. Sadi and A. J. Moghadam, 'Integration of vertical ground-coupled heat pump into a conventional natural gas pressure drop station: Energy, economic and CO2 emission assessment', *Energy*, vol. 112, pp. 998–1014, 1st Oct. 2016, ISSN: 0360-5442. DOI: 10.1016/j.energy.2016.06.100.

[14] M. Deymi-Dashtebayaz, M. Khorsand and H. R. Rahbari, 'Optimization of fuel consumption in natural gas city gate station based on gas hydrate temperature (case study: Abbas abad station)', *Energy & Environment*, vol. 30, no. 3, pp. 408–426, 1st May 2019, ISSN: 0958-305X. DOI: 10.1177/0958305X18793107.

[15] S. Englart, A. Jedlikowski, W. Cepiński and M. Badura, 'Renewable energy sources for gas preheating', *E3S Web of Conferences*, vol. 116, p. 00 019, 2019, ISSN: 2267-1242. DOI: 10.1051/e3sconf/201911600019.

[16] A. Jedlikowski, S. Englart, W. Cepiński, M. Badura and M. Ara Sayegh, 'Reducing energy consumption for electrical gas preheating processes', *Thermal Science and Engineering Progress*, vol. 19, p. 100 600, 1st Oct. 2020, ISSN: 2451-9049. DOI: 10.1016/j.tsep.2020.100600.

[17] A. Alizadeh, H. Ghadamian, M. Aminy, S. Hoseinzadeh, H. Khodayar Sahebi and A. Sohani, 'An experimental investigation on using heat pipe heat exchanger to improve energy performance in gas city gate station', *Energy*, vol. 252, p. 123 959, Aug. 2022, ISSN: 03605442. DOI: 10.1016/j.energy.2022.123959.

[18] D. Borelli, F. Devia, E. Lo Cascio and C. Schenone, 'Energy recovery from natural gas pressure reduction stations: Integration with low temperature heat sources', *Energy Conversion and Management*, vol. 159, pp. 274–283, 1st Mar. 2018, ISSN: 0196-8904. DOI: 10.1016/j.enconman.2017.12.084.

[19] E. Lo Cascio, D. Borelli, F. Devia and C. Schenone, 'Key performance indicators for integrated natural gas pressure reduction stations with energy recovery', *Energy Conversion and Management*, vol. 164, pp. 219–229, 15th May 2018, ISSN: 0196-8904. DOI: 10.1016/j.enconman.2018.02.089.

[20] P. Danieli, G. Carraro and A. Lazzaretto, 'Thermodynamic and economic feasibility of energy recovery from pressure reduction stations in natural gas distribution networks', *Energies*, vol. 13, no. 17, p. 4453, Jan. 2020, ISSN: 1996-1073. DOI: 10.3390/en13174453.

[21] P. Danieli, M. Masi, A. Lazzaretto, G. Carraro and G. Volpato, 'A smart energy recovery system to avoid preheating in gas grid pressure reduction stations', *Energies*, vol. 15, no. 1, p. 371, Jan. 2022, ISSN: 1996-1073. DOI: 10.3390/en15010371.

[22] S. Mohammad Ebrahimi Saryazdi, F. Rezaei, Y. Saboohi and F. Sassani, 'Multi-objective optimization of preheating system of natural gas pressure reduction station with turbo-expander through the application of waste heat recovery system', *Thermal Science and Engineering Progress*, vol. 38, p. 101 509, 1st Feb. 2023, ISSN: 2451-9049. DOI: 10.1016/j.tsep.2022.101509.

[23] A. Aramesh, N. Montazerin and A. Ahmadi, 'A general neural and fuzzy-neural algorithm for natural gas flow prediction in city gate stations', *Energy and Buildings*, vol. 72, pp. 73–79, Apr. 2014, ISSN: 03787788. DOI: 10.1016/j.enbuild.2013.12.020.

[24] M. Farzaneh-Gord, H. R. Rahbari, B. Mohseni-Gharyehsafa, A. Toikka and I. Zvereva, 'Machine learning methods for precise calculation of temperature drop during a throttling process', *Journal of Thermal Analysis and Calorimetry*, vol. 140, no. 6, pp. 2765–2778, Jun. 2020, ISSN: 1388-6150, 1588-2926. DOI: 10.1007/s10973-019-09029-3.

[25] M. B. Leo, A. Dutta, S. Farooq and I. A. Karimi, 'Simulation and health monitoring of a pressure regulating station', *Computers & Chemical Engineering*, vol. 139, p. 106 824, Aug. 2020, ISSN: 00981354. DOI: 10.1016/j.compchemeng.2020.106824.

[26] S. Sharma, M. B. Leo, S. Farooq and I. A. Karimi, 'Modelling and health monitoring of a pressure regulating system', in *Computer Aided Chemical Engineering*, ser. 31 European Symposium on Computer Aided Process Engineering, M. Türkay and R. Gani, Eds., vol. 50, Elsevier, 1st Jan. 2021, pp. 1383–1388. DOI: 10.1016/B978-0-323-88506-5.50213-8.

[27] L. Fan *et al.*, 'A deep reinforcement learning-based method for predictive management of demand response in natural gas pipeline networks', *Journal of Cleaner Production*, vol. 335, p. 130 274, 10th Feb. 2022, ISSN: 0959-6526. DOI: 10.1016/j.jclepro.2021.130274.

[28] C. Howard, P. Oosthuizen and B. Peppley, 'An investigation of the performance of a hybrid turboexpander-fuel cell system for power recovery at natural gas pressure reduction stations', *Applied Thermal Engineering*, Selected Papers from the 13th Conference on Process Integration, Modelling and Optimisation for Energy Saving and Pollution Reduction, vol. 31, no. 13, pp. 2165–2170, 1st Sep. 2011, ISSN: 1359-4311. DOI: 10.1016/j.applthermaleng.2011.04.023.

[29] M. Farzaneh-Gord, M. Khatib, M. Deymi-Dashtebayaz and M. Shahmardan, 'Producing electrical power in addition of heat in natural gas pressure drop stations by ICE', *Energy Exploration & Exploitation*, vol. 30, no. 4, pp. 567–587, 2012, ISSN: 0144-5987. DOI: 10.1260/0144-5987.30.4.567.

[30] S. Sanaye and A. M. Nasab, 'Modeling and optimizing a CHP system for natural gas pressure reduction plant', *Energy*, vol. 40, no. 1, pp. 358–369, Apr. 2012, ISSN: 0360-5442. DOI: 10.1016/j.energy.2012.01.060.

[31] I. Andrei, T. Valentin, T. Cristina and T. Niculae, 'Recovery of wasted mechanical energy from the reduction of natural gas pressure', *Procedia Engineering*, 24th DAAAM International Symposium on Intelligent Manufacturing and Automation, 2013, vol. 69, pp. 986–990, 1st Jan. 2014, ISSN: 1877-7058. DOI: 10.1016/j.proeng.2014.03.080.

[32] W. J. Kostowski, S. Usón, W. Stanek and P. Bargiel, 'Thermoecological cost of electricity production in the natural gas pressure reduction process', *Energy*, vol. 76, pp. 10–18, 1st Nov. 2014, ISSN: 0360-5442. DOI: 10.1016/j.energy.2014.01.045.

[33] M. A. Neseli, O. Ozgener and L. Ozgener, 'Energy and exergy analysis of electricity generation from natural gas pressure reducing stations', *Energy Conversion and Management*, vol. 93, pp. 109–120, 15th Mar. 2015, ISSN: 0196-8904. DOI: 10.1016/j.enconman.2015.01.011.

[34] V. Farzaneh-Kord *et al.*, 'Defining a technical criterion for economic justification of employing CHP technology in city gate stations', *Energy*, vol. 111, pp. 389–401, 15th Sep. 2016, ISSN: 0360-5442. DOI: 10.1016/j.energy.2016.05.122.

[35] A. Arabkoohsar, K. A. R. Ismail, L. Machado and R. N. N. Koury, 'Energy consumption minimization in an innovative hybrid power production station by employing PV and evacuated tube collector solar thermal systems', *Renewable Energy*, vol. 93, pp. 424–441, 1st Aug. 2016, ISSN: 0960-1481. DOI: 10.1016/j.renene.2016.03.003.

[36] R. Ghezelbash, M. Farzaneh-Gord and M. Sadi, 'Performance assessment of vortex tube and vertical ground heat exchanger in reducing fuel consumption of conventional pressure drop stations', *Applied Thermal Engineering*, vol. 102, pp. 213–226, 5th Jun. 2016, ISSN: 1359-4311. DOI: 10.1016/j.applthermaleng.2016.03.110.

[37] H. Tan, Q. Zhao, N. Sun and Y. Li, 'Proposal and design of a natural gas liquefaction process recovering the energy obtained from the pressure reducing stations of high-pressure pipelines', *Cryogenics*, vol. 80, pp. 82–90, 1st Dec. 2016, ISSN: 0011-2275. DOI: 10.1016/j.cryogenics.2016.09.010.

[38] A. Zabihi and M. Taghizadeh, 'Feasibility study on energy recovery at sari-akand city gate station using turboexpander', *Journal of Natural Gas Science and Engineering*, vol. 35, pp. 152–159, Sep. 2016, ISSN: 18755100. DOI: 10.1016/j.jngse.2016.08.054.

[39] P. Kolasiński, M. Pomorski, P. Błasiak and J. Rak, 'Use of rolling piston expanders for energy regeneration in natural gas pressure reduction stations—selected thermodynamic issues', *Applied Sciences*, vol. 7, no. 6, p. 535, Jun. 2017, ISSN: 2076-3417. DOI: 10.3390/app7060535.

[40] A. Diao, Y. Wang, Y. Guo and M. Feng, 'Development and application of screw expander in natural gas pressure energy recovery at city gas station', *Applied Thermal Engineering*, vol. 142, pp. 665–673, Sep. 2018, ISSN: 13594311. DOI: 10.1016/j.applthermaleng.2018.07.018.

[41] E. Lo Cascio, M. P. Von Friesen and C. Schenone, 'Optimal retrofitting of natural gas pressure reduction stations for energy recovery', *Energy*, vol. 153, pp. 387–399, 15th Jun. 2018, ISSN: 0360-5442. DOI: 10.1016/j.energy.2018.04.011.

[42] S. Khanmohammadi and M. Saadat-Targhi, 'Thermodynamic modeling and analysis of a novel heat recovery system in a natural gas city gate station', *Journal of Cleaner Production*, vol. 224, pp. 346–360, 1st Jul. 2019, ISSN: 0959-6526. DOI: 10.1016/j.jclepro.2019.03.167.

[43] S. Kuczynski, M. Laciak, A. Olijnyk, A. Szurlej and T. Wlodek, 'Techno-economic assessment of turboexpander application at natural gas regulation stations', *Energies*, vol. 12, no. 4, p. 755, 2nd Feb. 2019, ISSN: 1996-1073. DOI: 10.3390/en12040755.

[44] C. Li, S. Zheng, J. Li and Z. Zeng, 'Optimal design and thermo-economic analysis of an integrated power generation system in natural gas pressure reduction stations', *Energy Conversion and Management*, vol. 200, p. 112 079, 15th Nov. 2019, ISSN: 0196-8904. DOI: 10.1016/j.enconman.2019.112079.

[45] S. Yao, Y. Zhang, N. Deng, X. Yu and S. Dong, 'Performance research on a power generation system using twin-screw expanders for energy recovery at natural gas pressure reduction stations under off-design conditions', *Applied Energy*, vol. 236, pp. 1218–1230, 15th Feb. 2019, ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2018.12.039.

[46] A. Ebrahimi-Moghadam, A. J. Moghadam and M. Farzaneh-Gord, 'Comprehensive techno-economic and environmental sensitivity analysis and multi-objective optimization of a novel heat and power system for natural gas city gate stations', *Journal of Cleaner Production*, vol. 262, p. 121 261, 20th Jul. 2020, ISSN: 0959-6526. DOI: 10.1016/j.jclepro.2020.121261.

[47] S. Kowsari, M. Deymi-Dashtebayaz, K. Karbasi and H. Sheikhani, 'Optimal working conditions of various city gate stations for power and hydrogen production based on energy and eco-exergy analysis', *International Journal of Hydrogen Energy*, vol. 45, no. 43, pp. 22 513–22 533, 3rd Sep. 2020, ISSN: 0360-3199. DOI: 10.1016/j.ijhydene.2020.05.110.

[48] C.-h. Li, S.-y. Zheng, X.-y. Chen, J. Li and Z.-y. Zeng, 'Proposal and analysis of a coupled power generation system for natural gas pressure reduction stations', *Journal of Central South University*, vol. 27, no. 2, pp. 608–620, Feb. 2020, ISSN: 2095-2899. DOI: 10.1007/s11771-020-4320-3.

[49] M. Deymi-Dashtebayaz, D. Dadpour and J. Khadem, 'Using the potential of energy losses in gas pressure reduction stations for producing power and fresh water', *Desalination*, vol. 497, p. 114 763, 1st Jan. 2021, ISSN: 0011-9164. DOI: 10.1016/j.desal.2020.114763.

[50] S. M. E. Saryazdi, F. Rezaei and Y. Saboohi, 'Optimal detailed design and performance assessment of natural gas pressure reduction stations system equipped with variable inlet guide vane radial turbo-expander for energy recovery', *Journal of Natural Gas Science and Engineering*, vol. 96, p. 104 222, Dec. 2021, ISSN: 1875-5100. DOI: 10.1016/j.jngse.2021.104222.

[51] A. Ebrahimi-Moghadam and M. Farzaneh-Gord, 'Energy, exergy, and eco-environment modeling of proton exchange membrane electrolyzer coupled with power cycles: Application in natural gas pressure reduction stations', *Journal of Power Sources*, vol. 512, p. 230 490, 15th Nov. 2021, ISSN: 0378-7753. DOI: 10.1016/j.jpowsour.2021.230490.

[52] H. Golchoobian, S. Saedodin and B. Ghorbani, 'Exergetic and economic evaluation of a novel integrated system for trigeneration of power, refrigeration and freshwater using energy recovery in natural gas pressure reduction stations', *Journal of Thermal Analysis and Calorimetry*, vol. 145, no. 3, pp. 1467–1483, Aug. 2021, ISSN: 1388-6150. DOI: 10.1007/s10973-021-10607-7.

[53] A. H. S. Tabrizi, H. Niazmand, M. Farzaneh-Gord and A. Ebrahimi-Moghadam, 'Energy, exergy and economic analysis of utilizing the supercritical CO(2)recompression brayton cycle integrated with solar energy in natural gas city gate station', *Journal of Thermal Analysis and Calorimetry*, vol. 145, no. 3, pp. 973–991, Aug. 2021, ISSN: 1388-6150. DOI: 10.1007/s10973-020-10241-9.

[54] K. Wang and S.-s. Wang, 'Thermodynamic analysis of a comprehensive energy utilization system for natural gas pressure reduction stations based on allam cycle', *Applied Thermal Engineering*, vol. 205, p. 118 033, 25th Mar. 2022, ISSN: 1359-4311. DOI: 10.1016/j.applthermaleng.2021.118033.

[55] W. Xu, P. Zhao, F. Gou, A. Liu, W. Wu and J. Wang, 'Performance analysis of a power generation system for pressure energy recovery at natural gas city gate stations', *Applied Thermal Engineering*, vol. 213, p. 118 714, Aug. 2022, ISSN: 13594311. DOI: 10.1016/j.applthermaleng.2022.118714.

[56] A. Shahsavar, A. Jahangiri, A. Qatarani nejad, G. Ahmadi and A. Karamzadeh dizaji, 'Energy and exergy analysis and multi-objective optimization of using combined vortex tube-photovoltaic/thermal system in city gate stations', *Renewable Energy*, vol. 196, pp. 1017–1028, 1st Aug. 2022, ISSN: 0960-1481. DOI: 10.1016/j.renene.2022.07.057.

[57] E. Ashouri, F. Veysi, E. Shojaeizadeh and M. Asadi, 'The minimum gas temperature at the inlet of regulators in natural gas pressure reduction stations (CGS) for energy saving in water bath heaters', *Journal of Natural Gas Science and Engineering*, vol. 21, pp. 230–240, 1st Nov. 2014, ISSN: 1875-5100. DOI: 10.1016/j.jngse.2014.08.005.

[58] M. Farzaneh-Gord, A. Arabkoohsar, M. D. Dasht-bayaz, L. Machado and R. N. N. Koury, 'Energy and exergy analysis of natural gas pressure reduction points equipped with solar heat and controllable heaters', *Renewable Energy*, vol. 72, pp. 258–270, Dec. 2014, ISSN: 0960-1481. DOI: 10.1016/j.renene.2014.07.019.

[59] M. Farzaneh-Gord, R. Ghezelbash, A. Arabkoohsar, L. Pilevari, L. Machado and R. N. N. Koury, 'Employing geothermal heat exchanger in natural gas pressure drop station in order to decrease fuel consumption', *Energy*, vol. 83, pp. 164–176, 1st Apr. 2015, ISSN: 0360-5442. DOI: 10.1016/j.energy.2015.02.093.

[60] A. Zabihi and M. Taghizadeh, 'New energy-saving temperature controller for heater at natural gas gate station', *Journal of Natural Gas Science and Engineering*, vol. 27, pp. 1043–1049, Nov. 2015, ISSN: 1875-5100. DOI: 10.1016/j.jngse.2015.09.046.

[61] E. Lo Cascio, Z. Ma and C. Schenone, 'Performance assessment of a novel natural gas pressure reduction station equipped with parabolic trough solar collectors', *Renewable Energy*, vol. 128, pp. 177–187, Dec. 2018, ISSN: 0960-1481. DOI: 10.1016/j.renene.2018.05.058.

[62] M. Moghimi, M. Hosseinnia and M. Emadi, 'Energy and environmental analysis of a natural gas pressure reduction station equipped with turboexpander, solar collector, and storage tank', *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 40, no. 9, p. 434, Sep. 2018, ISSN: 1678-5878. DOI: 10.1007/s40430-018-1345-8.

[63] M. Naderi, G. Ahmadi, M. Zarringhalam, O. Akbari and E. Khalili, 'Application of water reheating system for waste heat recovery in NG pressure reduction stations, with experimental verification', *Energy*, vol. 162, pp. 1183–1192, 1st Nov. 2018, ISSN: 0360-5442. DOI: 10.1016/j.energy.2018.08.111.

[64] M. Khosravi, A. Arabkoohsar, A. S. Alsagri and M. Sheikholeslami, 'Improving thermal performance of water bath heaters in natural gas pressure drop stations', *Applied Thermal Engineering*, vol. 159, p. 113 829, Aug. 2019, ISSN: 1359-4311. DOI: 10.1016/j.applthermaleng.2019.113829.

[65] Y. Noorollahi, G. Ghasemi, F. Kowsary, S. Roumi and S. Jalilinasrabady, 'Modelling of heat supply for natural gas pressure reduction station using geothermal energy', *International Journal of Sustainable Energy*, vol. 38, no. 8, pp. 773–793, 2019, ISSN: 1478-6451. DOI: 10.1080/14786451.2019.1585434.

[66] M. Olfati, M. Bahiraei and F. Veysi, 'A novel modification on preheating process of natural gas in pressure reduction stations to improve energy consumption, exergy destruction and CO2 emission: Preheating based on real demand', *Energy*, vol. 173, pp. 598–609, 15th Apr. 2019, ISSN: 0360-5442. DOI: 10.1016/j.energy.2019.02.090.

[67] A. Sircar and K. Yadav, 'Optimization of city gas network: A case study from gujarat, india', *Sn Applied Sciences*, vol. 1, no. 7, p. 769, Jul. 2019, ISSN: 2523-3963. DOI: 10.1007/s42452-019-0755-2.

[68] A. Ebrahimi-Moghadam, M. Deymi-Dashtebayaz, H. Jafari and A. Niazmand, 'Energetic, exergetic, environmental and economic assessment of a novel control system for indirect heaters in natural gas city gate stations', *Journal of Thermal Analysis and Calorimetry*, vol. 141, no. 6, pp. 2573–2588, Sep. 2020, ISSN: 1388-6150. DOI: 10.1007/s10973-020-09413-4.

[69] S. A. Mostafavi and M. Shirazi, 'Thermal modeling of indirect water heater in city gate station of natural gas to evaluate efficiency and fuel consumption', *Energy*, vol. 212, p. 118 390, 1st Dec. 2020, ISSN: 0360-5442. DOI: 10.1016/j.energy.2020.118390.

[70] M. Olfati, M. Bahiraei, S. Nazari and F. Veysi, 'A comprehensive assessment of low-temperature preheating process in natural gas pressure reduction stations to better benefit from solar energy', *Energy*, vol. 209, p. 118 430, Oct. 2020, ISSN: 03605442. DOI: 10.1016/j.energy.2020.118430.

[71] Y. Sheikhnejad, J. Simoes and N. Martins, 'Energy harvesting by a novel substitution for expansion valves: Special focus on city gate stations of high-pressure natural gas pipelines', *Energies*, vol. 13, no. 4, p. 956, Feb. 2020. DOI: 10.3390/en13040956.

[72] P. Soleimani, M. Khoshvaght-Aliabadi, H. Rashidi and H. Bahmanpour, 'Performance enhancement of water bath heater at natural gas city gate station using twisted tubes', *Chinese Journal of Chemical Engineering*, vol. 28, no. 1, pp. 165–179, Jan. 2020, ISSN: 1004-9541. DOI: 10.1016/j.cjche.2019.03.018.

[73] K. Esfandiari, P. Soleimani and E. Hakimi, 'Investigation of operational scenarios to mitigate CO2 emission and natural gas consumption in city gate stations (CGSs)', *Process Integration and Optimization for Sustainability*, vol. 5, no. 1, pp. 63–73, Mar. 2021, ISSN: 2509-4238. DOI: 10.1007/s41660-020-00144-5.

[74] R. Alayi, M. Jahangiri and A. Najafi, 'Energy analysis of vacuum tube collector system to supply the required heat gas pressure reduction station', *International Journal of Low-Carbon Technologies*, vol. 16, no. 4, pp. 1391–1396, Dec. 2021, ISSN: 1748-1317. DOI: 10.1093/ijlct/ctab069.

[75] D. Shafiei, S. A. Mostafavi, S. J. Mehrabadi and D. Toghraie, 'Analysis of the effects of forced convective heat transfer to reduce the efficiency of heaters of gas pressure reducing stations', *Case Studies in Thermal Engineering*, vol. 38, p. 102 356, Oct. 2022, ISSN: 2214157X. DOI: 10.1016/j.csite.2022.102356.

[76] R. Nourian and S. M. Mousavi, 'Design and implementation of an expert system for periodic and emergency control under uncertainty: A case study of city gate stations', *Journal of Natural Gas Science and Engineering*, vol. 66, pp. 306–315, Jun. 2019, ISSN: 1875-5100. DOI: 10.1016/j.jngse.2019.04.007.

[77] R. Nourian, S. M. Mousavi and S. Raissi, 'A fuzzy expert system for mitigation of risks and effective control of gas pressure reduction stations with a real application', *Journal of Loss Prevention in the Process Industries*, vol. 59, pp. 77–90, 1st May 2019, ISSN: 0950-4230. DOI: 10.1016/j.jlp.2019.03.003.

[78] L. Zuo, S. Zhao, Y. Ma, F. Jiang and Y. Zu, 'Natural gas hydrate prediction and prevention methods of city gate stations', *Mathematical Problems in Engineering*, vol. 2021, p. 5 977 460, 14th Jul. 2021, ISSN: 1024-123X. DOI: 10.1155/2021/5977460.

[79] P. Dadkani, E. Noorzai, A. Ghanbari and A. Gharib, 'Risk analysis of gas leakage in gas pressure reduction station and its consequences: A case study for zahedan', *Heliyon*, vol. 7, no. 5, e06911, May 2021. DOI: 10.1016/j.heliyon.2021.e06911.

[80] W. Panek and T. Włodek, 'Natural gas consumption forecasting based on the variability of external meteorological factors using machine learning algorithms', *Energies*, vol. 15, no. 1, p. 348, Jan. 2022, Number: 1 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 1996-1073. DOI: 10.3390/en15010348.

[81] F. Pedregosa *et al.*, 'Scikit-learn: Machine learning in python', *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011, ISSN: 1533-7928.

[82] I. H. Bell, J. Wronski, S. Quoilin and V. Lemort, 'Pure and pseudo-pure fluid thermophysical property evaluation and the open-source thermophysical property library CoolProp', *Industrial & Engineering Chemistry Research*, vol. 53, no. 6, pp. 2498–2508, 12th Feb. 2014, ISSN: 0888-5885, 1520-5045. DOI: 10.1021/ie4033999.

[83] T. Huld, R. Müller and A. Gambardella, 'A new solar radiation database for estimating PV performance in europe and africa', *Solar Energy*, vol. 86, no. 6, pp. 1803–1815, 1st Jun. 2012, ISSN: 0038-092X. DOI: 10.1016/j.solener.2012.03.006.

[84] sherena.johnson@nist.gov. 'REFPROP', NIST. Last Modified: 2022-01-27T11:03-05:00. (18th Apr. 2013), [Online]. Available: https://www.nist.gov/srd/refprop (visited on 04/06/2022).

[85] S. Quoilin, K. Kavvadias, A. Mercier, I. Pappone and A. Zucker, 'Quantifying self-consumption linked to solar home battery systems: Statistical analysis and economic assessment', *Applied Energy*, vol. 182, pp. 58–67, 15th Nov. 2016, ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2016.08.077.

[86] S. Kozarcanin, R. Hanna, I. Staffell, R. Gross and G. B. Andresen, 'Impact of climate change on the cost-optimal mix of decentralised heat pump and gas boiler technologies in europe', *Energy Policy*, vol. 140, p. 111 386, 1st May 2020, ISSN: 0301-4215. DOI: 10.1016/j.enpol.2020.111386.

[87] E. Venegas-Reyes, N. Ortega-Avila, N. A. Rodríguez-Muñoz, M. Nájera-Trejo, I. R. Martín-Domínguez and J. Ibarra-Bahena, 'Parametric methodology to optimize the sizing of solar collector fields in series-parallel arrays', *Processes*, vol. 7, no. 5, p. 294, May 2019, Number: 5 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 2227-9717. DOI: 10.3390/pr7050294.

[88] 'GSE'. (), [Online]. Available: https://www.gse.it/.

[89] '2020 gas decarbonisation pathways study', 2020.

[90] L. M. Pastore, G. Lo Basso and L. de Santoli, 'Towards a dramatic reduction in the european natural gas consumption: Italy as a case study', *Journal of Cleaner Production*, vol. 369, p. 133 377, 1st Oct. 2022, ISSN: 0959-6526. DOI: 10.1016/j.jclepro.2022.133377.

[91] D. Fischer, K. B. Lindberg, H. Madani and C. Wittwer, 'Impact of PV and variable prices on optimal system sizing for heat pumps and thermal storage', *Energy and Buildings*, vol. 128, pp. 723–733, Sep. 2016, ISSN: 03787788. DOI: 10.1016/j.enbuild.2016.07.008.

[92] 'GME - gestore dei mercati energetici SpA'. (), [Online]. Available: https://www.mercatoelettrico.org/it/ (visited on 25/05/2023).

[93] W. Zhang, H. Li, Y. Li, H. Liu, Y. Chen and X. Ding, 'Application of deep learning algorithms in geotechnical engineering: A short critical review', *Artificial Intelligence Review*, vol. 54, no. 8, pp. 5633–5673, 1st Dec. 2021, ISSN: 1573-7462. DOI: 10.1007/s10462-021-09967-1.

[94] 'Machine learning mastery', MachineLearningMastery.com. (), [Online]. Available: https://machinelearningmastery.com/ (visited on 27/05/2023).

[95] 'Towards data science', Towards Data Science. (), [Online]. Available: https://towardsdatascience.com (visited on 27/05/2023).

[96] V. S. Puranik, 'CUSUM quality control chart for monitoring energy use performance', in *2007 IEEE International Conference on Industrial Engineering and Engineering Management*, ISSN: 2157-362X, Dec. 2007, pp. 1231–1235. DOI: 10.1109/IEEM.2007.4419388.

[97] R. Nouvel, M. Cotrado Sehgelmeble and D. Pietruschka, *European Mapping of Seasonal Performances of Air-source and Geothermal Heat Pumps for Residential Applications*. 9th Sep. 2015.