ENERGY EFFICIENCY IN INDUSTRIAL FACILITIES

Improvements on energy transformation and data analysis

Ph.D. student: Giulio Vialetto

Supervisor: Prof. Marco Noro

Fatti non foste a viver come bruti, ma per seguir virtute e canoscenza.

(Dante, "*La divina commedia*", Canto XVI)

Think of the seed of you were not born to live as brutes, but to follow virtue and knowledge.

ACKNOWLEDGMENT

During these three years of Ph.D. in many moments it was hard to continue, and it seems difficult to finish it: I would like to thank all my parent, relatives and friends who took care of me. I am grateful to prof Marco Noro who supported me during all the period of the Ph.D., for the continuous advices to keep the focus even if the research field was so vast and it was easy to get lost. A special thanks to Corà Domenico e Figli S.p.A., in particular to Stefano Corà, Mauro Milanesio and Maurizio Rigotto that provided useful dataset of energy demand that allowed me to test machine learning and data analytics methods developed during my Ph.D. and presented in this thesis. A special thanks also to Mosaico s.r.l. (part of Cartiere Burgo S.p.A.), Lugo di Vicenza plant who provided an interesting case study with a dataset regarding energy generation into industrial facility. I would thank Dr Marvin Mikael Rokni, associate professor at Technical University of Denmark, and Dr. Paolo Colbertaldo, research fellow at Politecnico di Milano (Italy), for the collaboration on SOFC and RSOC and the articles we published together.

ABSTRACT

During these years of my Ph.D. studies the main aim of the research work was to improve the efficiency on energy generation into industrial facilities. Novelties are proposed both on the devices used for energy generation and on energy consumption data analytics. In the first part of the thesis, Solid Oxide Fuel Cell (SOFC) and Reversible Solid Oxide Cell (RSOC) are proposed: these technologies have many advantages such as high efficiency on energy generation, heat available at high temperature, and modularity.

A new heat recovery for a modular micro-cogeneration system based on SOFC is presented with the main goal of improving the efficiency of an air source heat pump with unused heat of fuel cell exhausted gases. The novelty of the system proposed is that exhaust gases after the fuel cell are firstly used to heat water and/or used to produce steam, then they are mixed with the external air to feed the evaporator of the heat pump with the aim of increasing energy efficiency of the latter. This system configuration decreases the possibility of freezing of the evaporator as well, which is one of the drawbacks for air source heat pump in climates where temperature close to 0 °C and high humidity could occur. Results show that the performance of the air source heat pump increases considerably during cold season for climates with high relative humidity and for users with high electric power demand.

As previously cited, not only SOFC but also RSOC are deeply analysed in the thesis to define innovative energy generation system with the possibility of varying H/P ratio to match energy generation and demand in order to avoid mismatching and, consequently, integration system with a lower system. The aim is to define a modular system where each RSOC module can be switched between energy generation mode (fuel consumption to produce electricity and heat) and energy consumption (electricity and heat are consumed to produce hydrogen, working as Solid Oxide Electrolysis Cells) to vary overall H/P of the overall system. Hydrogen is a sub-product of the system and can be used for many purposes such as fuel and/or for transport sector. Then a re-vamping of the energy generation system of a paper mill by means of RSOCS is proposed and analysed: a real industrial facility, based in Italy with a production capacity of 60000 t/y of paper, is used as case study. Even if the complexity of the system increases, results show that saving between 2% and 6% occurs. Hydrogen generation is assessed, comparing the RSOC integrated system with PEM electrolysis, in terms of both primary energy and economics. Results exhibit significant primary energy and good economic performance on hydrogen production with the novel system proposed.

In the thesis novelties are proposed not only on energy system “hardware” (component for energy generation) but also on “software”. In the second part of the thesis, artificial intelligence and machine learning methods are analysed to perform analytics on energy consumption data and consequently to improve performances on energy generation and operation strategy.

A study on how cluster analysis could be applied to analyse energy demand data is depicted. The aim of the method is to design cogeneration systems that suit more efficiently energy demand

profiles, choosing the correct type of cogeneration technology, operation strategy and, if they are necessary, energy storages. A case study of a wood industry that requires low temperature heat to dry wood into steam-powered kilns that already uses cogeneration is proposed to apply the methodology in order to design and measure improvements. An alternative cogeneration system is designed and proposed, thermodynamics benchmarks are defined to evaluate differences between as-is and alternative scenarios. Results show that the proposed innovative method allows to choose a more suitable cogeneration technology compared to the adopted one, giving suggestions on the operation strategy in order to decrease energy losses and, consequently, primary energy consumption.

Finally, clustering is suggested for short-term forecasting of energy demand in industrial facilities. A model based on clustering and kNN is proposed to find similar pattern of consumption, to identify average consumption profiles, and then to use them to forecast consumption data. Novelties on model parameters definition such as data normalisation and clustering hyperparameters are presented to improve its accuracy. The model is then applied to the energy dataset of the wood industry previously cited. Analysis on the parameters and the results of the model are performed, showing a forecast of electricity demand with an error of 3%.

PREFACE

In the last centuries humankind improved its life condition, expectation and wealth. At the end of the 18th century, the first industrial revolution occurred, and the economy started a transition from being based on agriculture to being based on the industry thus enhancing life conditions. From 18th to 21st century, other industrial revolutions occurred (Figure P. 1). The first industrial revolution was mainly an introduction of mechanisation of industrial processes. The second industrial revolution improved the efficiency of industrial processes by introducing mass production and labour division. Successively, with the third industrial revolution electronics and information technologies such as computer, programmable logic controller (PLC) and robot were introduced. More recently, industrial internet of thing (IIOT), big data, machine learning and artificial intelligence (AI) were introduced by the fourth industrial revolution.

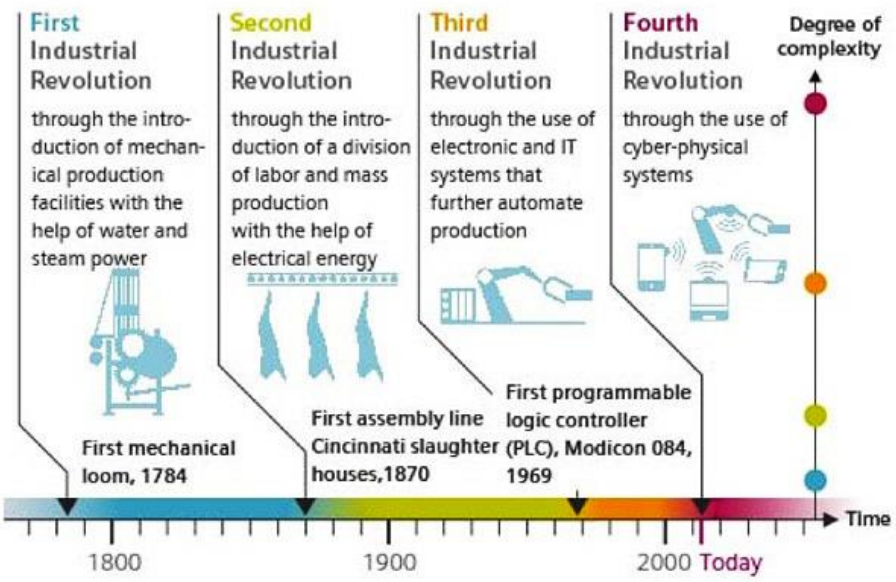


Figure P. 1. Timeline of the industrial revolutions (from www.manufacturing-operations-management.com)

The increase in energy consumption is one of the main consequences of these revolutions, as it is related to the worldwide growth of the industrial sector. Energy is necessary for industry, and it is used for many purposes: for example, it could be used for lighting, to produce heat and/or cooling, as mechanical energy by motors thus reducing human manual labour, or for chemical processes as electroplating. During the first industrial revolution, the primary energy source used was coal, whereas oil was mainly used during the second. Electricity was introduced during the second industrial revolution increasing the overall energy consumption: even if it is not a primary energy source but an energy vector, countries such as Italy (that is poor of coal and oil) were able to produce energy using renewable energies like hydropower (Figure 2). Industrial processes used electricity to improve power and efficiency: for example, during the first industrial revolution only steam motors

were available to produce mechanical energy, whereas electrical motors with higher versatility and efficiency on energy conversion were successively developed.

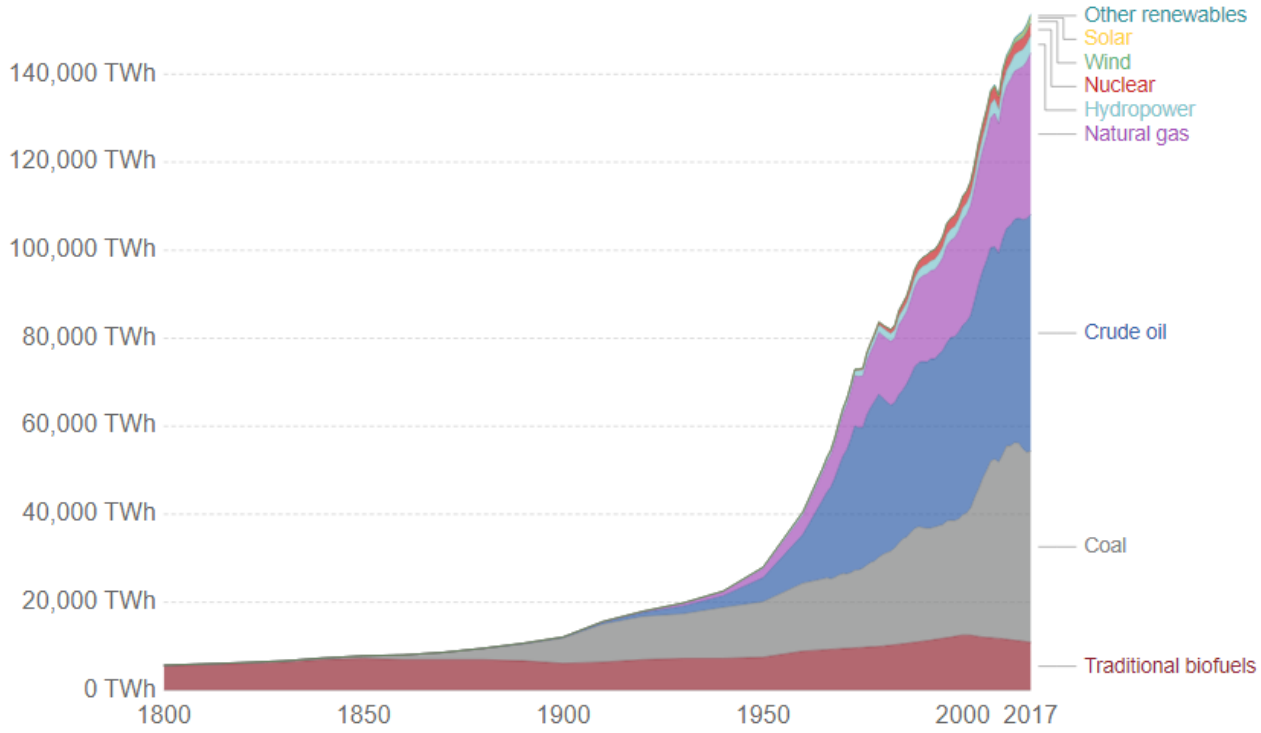


Figure P. 2. World historical energy consumption divided by energy source (from https://ourworldindata.org/energy-production-and-changing-energy-sources)

Meanwhile the industrial sector increases its share on the world economy and, consequently, also on energy consumption and the use of primary energy sources such as coal, oil, natural gas or hydrocarbons has increased CO_2 concentration into the atmosphere. Nowadays, CO_2 has reached the highest concentration in the last 800 000 years (Figure P. 3 and Figure P. 4).

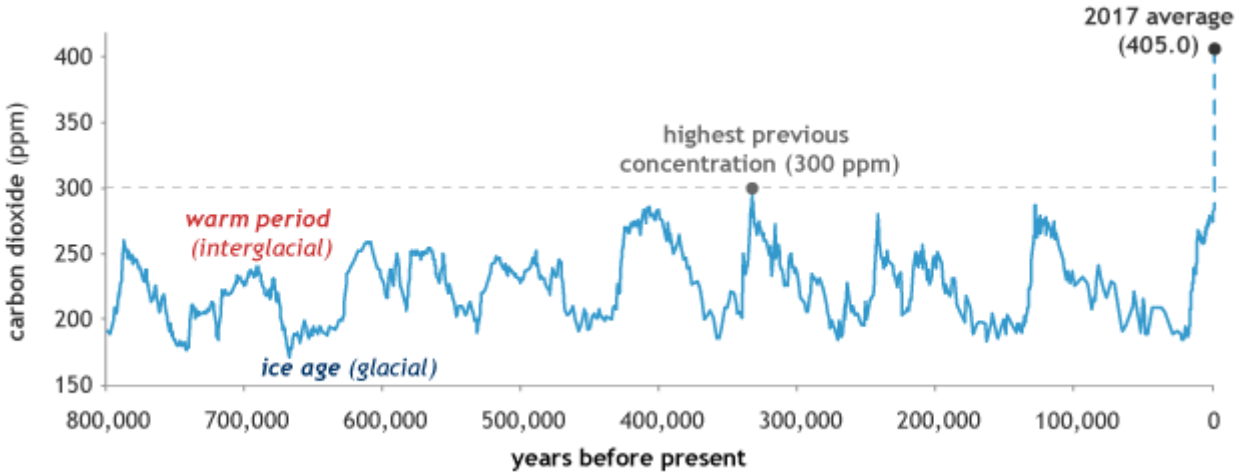


Figure P. 3. World historical CO_2 level into the atmosphere (from www.climate.gov/news-features/understanding-climate/climate-change-atmospheric-carbon-dioxide)

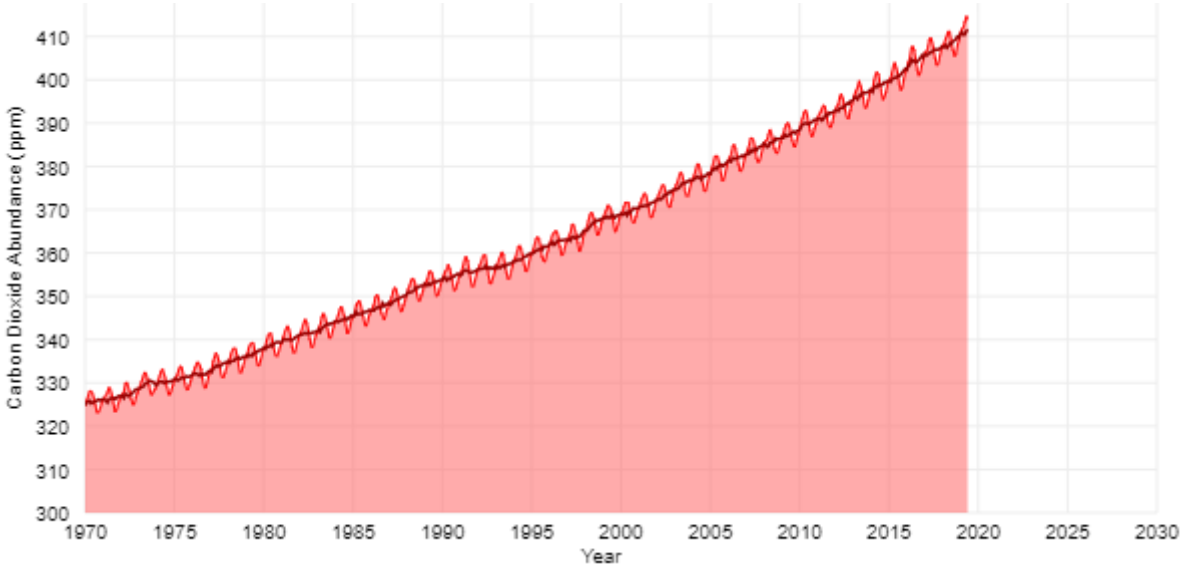


Figure P. 4. World CO_2 level into the atmosphere from 1970 to 2017 (from www.climate.gov/news-features/understanding-climate/climate-change-atmospheric-carbon-dioxide)

The scientific community is worried about the consequences of the increasing of CO_2 level, one of the most known being global warming. CO_2 is one of the gases related to greenhouse effect: if its concentration increases, the atmosphere decreases its transparency with respect to infrared radiation from the Earth surface to the space. Earth has a natural greenhouse effect thanks to the natural presence of greenhouse gases that allow a temperature suitable for life. However, during the last centuries anthropic activity has been increasing continuously the concentration of greenhouse gases (not only CO_2 but also CH_4 and N_2O), and consequently the global average temperature has increased too (Figure P. 5).

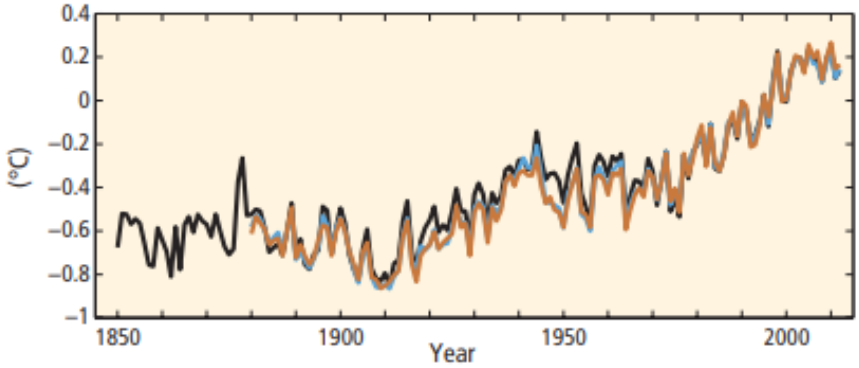


Figure P. 5. Globally averaged combined land and ocean surface temperature anomaly, different colours refer to different temperature dataset (from Climate Change 2014 Synthesis Report Summary for Policymakers – IPCC, www.ipcc.ch)

The scientific community has already warned policymakers about the global warming effects on atmosphere temperature and, then, on the whole climate: sea level increase, change on

precipitation and modification on the climate are expected effects. The Intergovernmental Panel on Climate Change (IPCC) of United Nation published the Climate Change 2014 Synthesis Report Summary for Policymakers, where different scenarios based on global emission of CO_2 are proposed. The increase of temperature by the end of the 21st century is likely to be between 0.3 °C and 1.7 °C when considering the lower emission scenario (RCP2.6), and between 2.6 °C and 4.8 °C with the higher emission scenario (RCP8.5) (Figure P. 6).

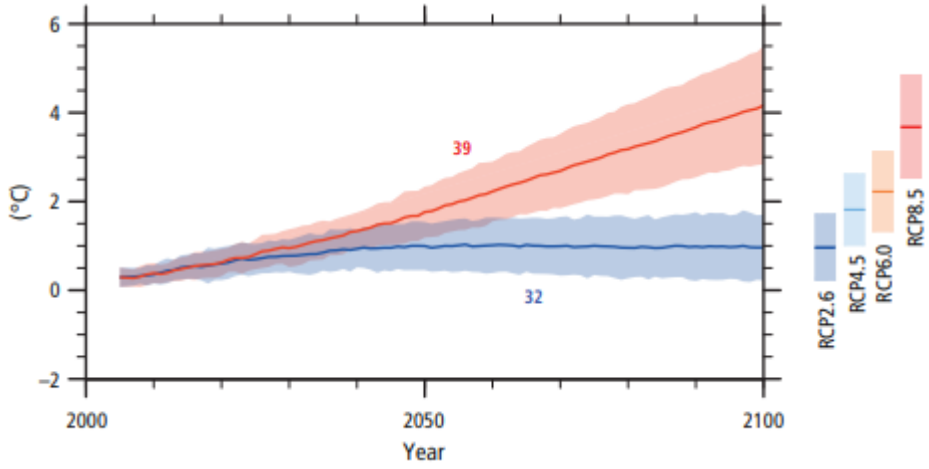


Figure P. 6. Global average surface temperature change, 2081-2100 mean temperature versus 1986-2005, red projections and time series are related to the worst scenario (RCP8.5), blue ones to the lower scenario (RCP2.6), intermediate scenarios are also presented (RCP4.5 and RCP6.0) (from Climate Change 2014 Synthesis Report Summary for Policymakers – IPCC, www.ipcc.ch)

Global sea levels are likely to increase with a higher rate under RCP8.5 model (worst scenario) than RCP2.6 (Figure P. 7). Also changing on precipitation are expected (Figure P. 8).

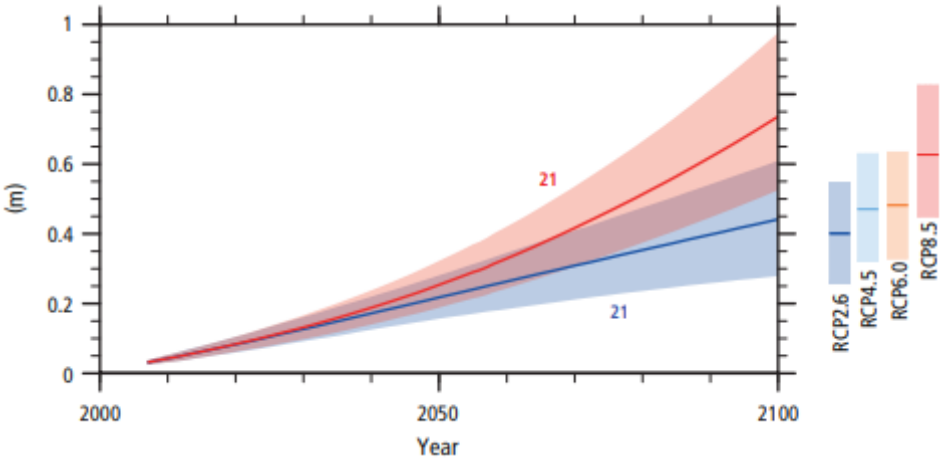


Figure P. 7. Global mean sea level rise, 2081-2100 mean sea level versus 1986-2005, red projections and time series are related to the worst scenario (RCP8.5), blue ones to the lower

scenario (RCP2.6), intermediate scenarios are also presented (RCP4.5 and RCP6.0) (from Climate Change 2014 Synthesis Report Summary for Policymakers – IPCC, www.ipcc.ch)

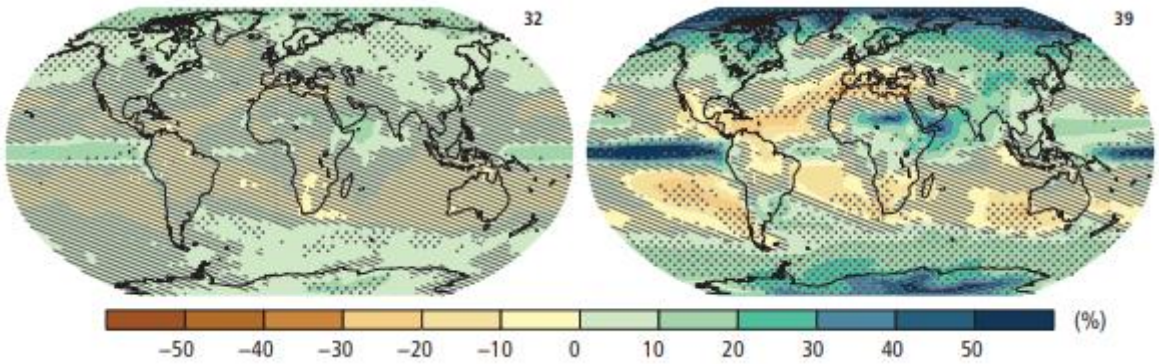


Figure P. 8. Change in average precipitation, 2081-2100 average precipitation (right) versus 1986-2005 average (left). Data refers to the average of the models previous cited (from Climate Change 2014 Synthesis Report Summary for Policymakers – IPCC, www.ipcc.ch)

Industry has an important role in CO_2 emissions as it is one of the largest energy consuming sector compared to transport, residential and tertiary. According to the International Energy Agency (IEA), the share of the total final energy consumption in 2016 was 31.7% for industry, 31.6% for transport and 36.7% for other uses (mainly residential and tertiary) (Figure P. 9).

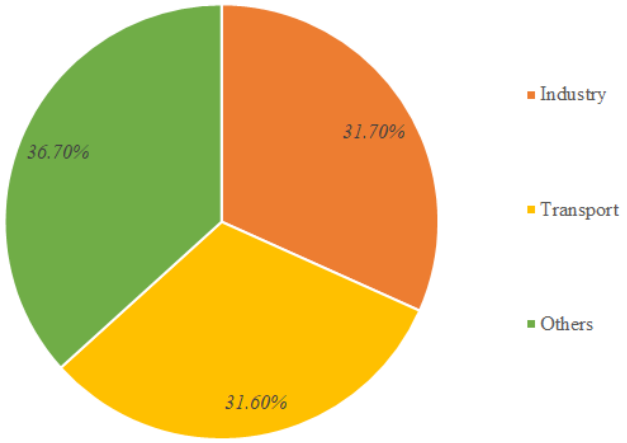


Figure P. 9. World energy consumption by sector in 2016 (data from International Energy Agency – IEA Key World Energy Statistics)

Considering the first five countries for total energy consumption (Japan, Russia, India, United States and China), it is possible to appreciate how much industry has an impact on the energy consumption: 34.54% of 4824 Mtoe of the total final energy consumption is due to the industry sector.

China features the highest share: industry sector consumes 50% on final energy consumption (Figure P. 10).

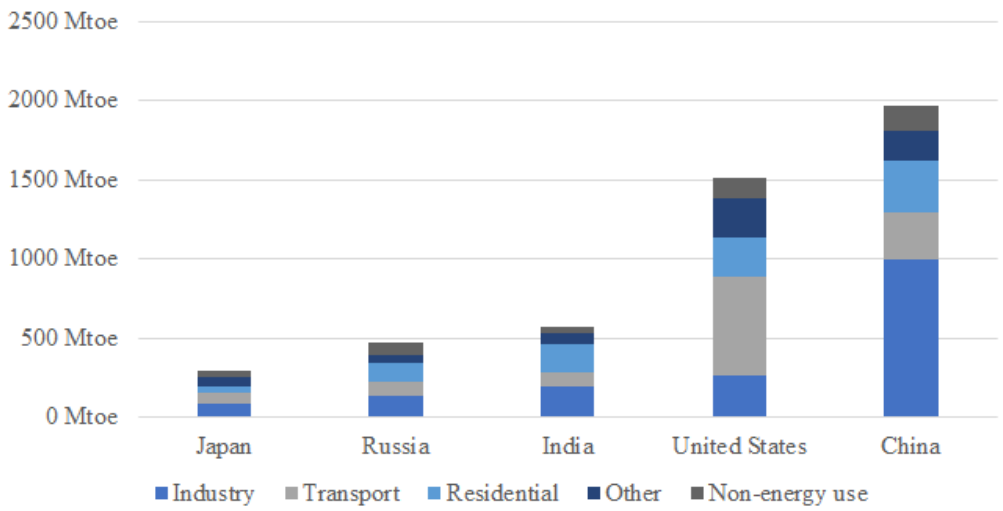


Figure P. 10. Top five final energy consumers divided by sector in 2016 (data from International Energy Agency – IEA Key World Energy Statistics)

Figure P. 11 shows the world primary energy sources: more than 80% are hydrocarbons (oil or natural gas) and coal. These energy sources are strictly related to global warming effect because of the CO_2 emissions during the combustion process. As the industrial sector has the highest share of energy consumption compared to transport and other sectors, it has a high responsibility on greenhouse gases emissions. As a matter of fact, reduction of the consumption of primary energy sources, increase of energy efficiency and use of renewables are key drivers for the industry to decrease greenhouse gases emissions as suggested by IPCC.

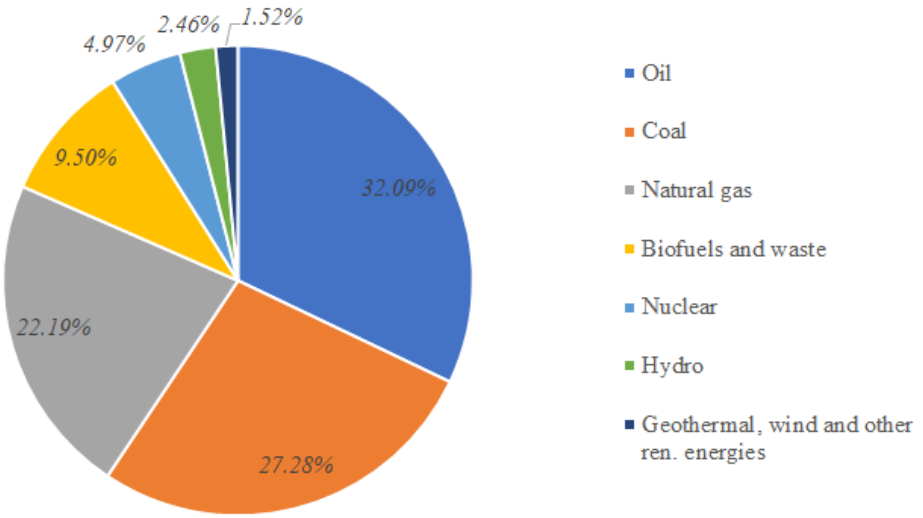


Figure P. 11. World primary energy supply by source in 2016 (data from International Energy Agency – IEA Key World Energy Statistics)

Italy is considered one of the most industrialised countries in the world. Italian industrial sector has also a unique peculiarity in the world: industrial districts. An industrial district is a cluster of industries (usually small-medium enterprises, SME), working on the same or similar businesses, that are geographically located in a (relative) small area. This can bring many advantages: according to the latest report of Intesa San Paolo bank, industrial districts have a shorter supply chain, a higher innovation rate and consequently they usually have higher profitability (*return of equity - ROE*) and labour productivity (*Economia e finanza dei distretti industriali – Rapporto annuale – n. 11 Dicembre 2018*). The Italian National Institute of Statistics (ISTAT) defined 141 different districts during its last national industrial census (Figure P. 12). The majority (38, about 27% of the total) are classified as mechanical industry districts, then textile and clothes (32, about 22.7 %), furniture and household goods (24, 17% of the total), leather and tannery industry (17, 12.1%), food industry (15, about 10.6 %). The last 10% are classified as chemical industry (5), jewellery, musical instruments (4), metallurgical industry (4) and pulp and paper industry (2).

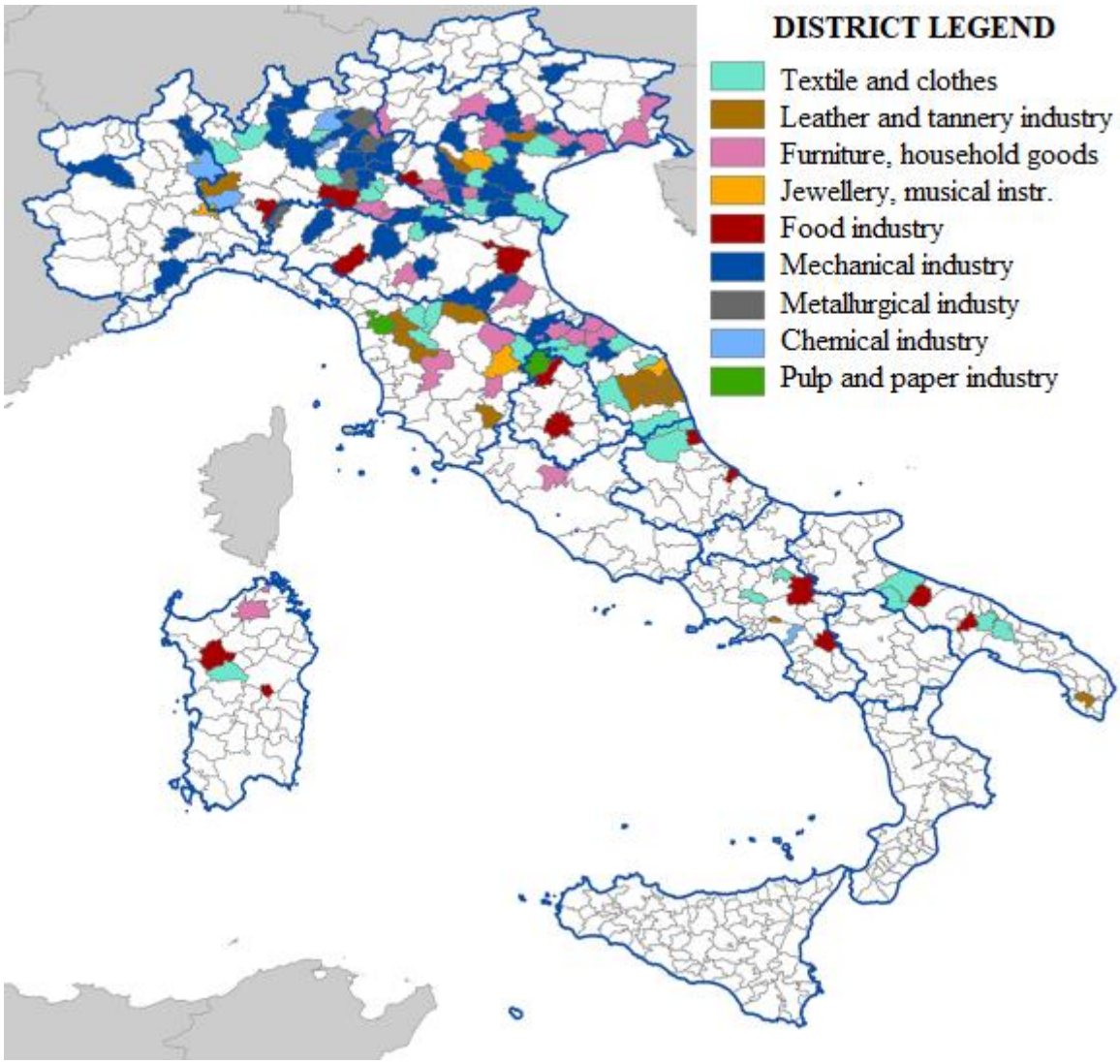


Figure P. 12. Italian industrial districts, 2011 census (Data from ISTAT, I distretti industriali Anno 2011, https://www.istat.it/it/archivio/150320)

Moreover, Italian industrial districts can take advantage of smart and modular energy systems to suit energy demand of the industrial clusters that have homogeneous sectors and, consequently, similar average energy consumptions.

Not only cogeneration and/or improvements on energy generation could be used to increase efficiency on energy systems: big data and, consequently, machine learning and artificial intelligence are trend topics on *smart energy systems*. Information technologies (IT) are expected to revolutionise energy sector (Figure P. 13). An example of this transformation could be smart grids. Usually, energy grids (and electricity grid in particular) are designed with a centralized unidirectional generation site and distribution: there are few (generally big) generators, and many (big or small) users. Smart grids, instead, have de-centralised generators, and each component of the grid could switch from consumer to producer. In this case, IT are necessary to manage the grid, for example load balancing is easier with centralised generators than in smart grids where small and de-centralised generators must be coordinated.

An increase of renewable energy plants such as solar or wind should be preferred to decrease greenhouses emissions. One of the main issues is the planning and management of renewables, as they are highly affected by weather conditions. As a matter of fact, these energy sources have been increasing their share on overall energy production in the last years, and IT tools must be developed and implemented. *Kaile Zhou et al.* have analysed the use of big data and machine learning in energy sector and their impact for each component of the energy system, from power generation to energy consumption (Figure P. 13).

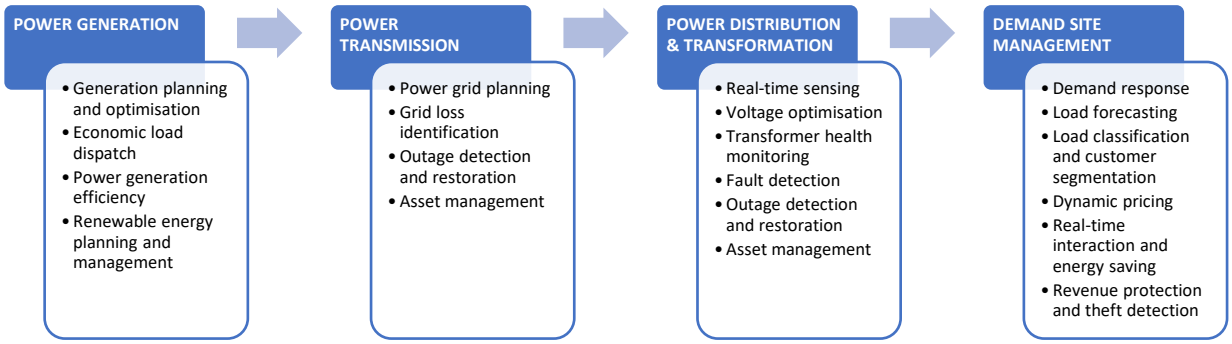


Figure P. 13. Big data support on smart energy system (Kaile Zhou et al., Big data driven smart energy system management: From big data to big insights, Renewable and Suitable Energy Reviews)

In this thesis, the topic of energy efficiency related to industrial facilities is tackled. As previously mentioned, the industrial sector has a high energy demand with consequences on greenhouse gases emissions. The aim of the work is to increase efficiency on energy generation and conversion into industrial facility: lower consumptions (and emission and costs) are then expected. The main topics stress on energy generation and conversion instead of improving energy saving of

industrial processes: energy generation is similar between different industrial sectors, but industrial processes change and could be completely different between each other. Moreover, if improvements are achieved on energy generation, they could be introduced in many sectors. Enhancements on industrial processes require, instead, a deep knowledge of each process and, if improvements could be achieved, they could not be shared with other sectors. In this thesis, improvements are proposed by increasing efficiency on energy generation by means of alternative energy generation systems and big data methods.

Firstly, alternative energy systems are proposed to increase efficiency on energy generation, with a focus on polygeneration systems where different types of energy are produced simultaneously. An industrial facility for example could require electricity and heat at different temperatures, low level (lower than 50 °C) and medium level (up to 100 °C). Polygeneration is expected to increase efficiency compared to separate energy generation systems: a common example is cogeneration. Fuel cells, in particular solid oxide fuel cells (SOFC), are proposed due to their advantages: modularity and flexibility, high efficiency on energy conversion, waste heat available at high temperature (600 °C – 1000 °C) and flexibility on fuel.

In the first part of the thesis, an analysis of solid oxide fuel cells is performed. This type of fuel cell can convert the chemical energy of hydrocarbons (for example propane, natural gas, etc) and ammonia directly into electricity and heat. SOFC work at high temperature (600 °C – 1000 °C) with high efficiency both on electricity production and overall energy conversion. Innovative energy systems based on SOFC are proposed to increase efficiency on energy production, firstly with a modular cogeneration system where air source heat pump improves heat production using SOFC exhausted gases increasing its coefficient of performance. The proposed system could produce electricity and heat at different levels: a higher level using exhausted gases of SOFC (around 1000 °C), and a lower level (about 50 °C) from air source heat pump. Applications in industrial facilities are possible as some processes could require heat at different temperatures and/or combining industrial energy demand with district heating.

A polygeneration system with hydrogen production based on SOFC and solid oxide electrolyser is then proposed: the aim is to vary heat to power ratio on energy generation with hydrogen production to match heat to power user request with a high efficiency both on energy production (using SOFC cells) and on hydrogen production (using solid oxide electrolyser). A paper mill is analysed and proposed to introduce this innovative energy system to increase efficiency on energy generation, dismissing a part of an old steam cogeneration system meanwhile decreasing energy costs. Simulations prove that primary energy saving could be achieved also with hydrogen production.

In the second part of the thesis machine learning, big data and artificial intelligence are explored. The starting point is related to the huge amount of data that are usually collected into Enterprise Resource Planning (ERP) system. Such data are not often used to query and perform analytic studies on it. These methods could provide analytics on collected data, allowing improvements on energy efficiency of the industrial facility. Moreover, they can give the opportunity

to forecast energy consumption and continuous monitoring of the correct operation and/or need of maintenance of the production plants.

The aim of this part of the thesis is to use these methods in order to increase the efficiency of energy systems. The main idea is that improvements on energy generation technologies ("hardware") are only a "part of the effort" to achieve energy savings in industry. Data analysis, suitable operation strategies, smart energy system controls and load forecasting ("software") can be developed and applied as well. As a matter of fact, if data analytics is not applied to increase the matching between energy demand and production, or load forecasting is not used to correctly choose the size of the generation plant and the storage, energy dissipation could occur even if energy systems with high performance (such as SOFC) are used. Improvements on the previous cited "software" are as important as improvements on the "hardware" of the energy system. Even if different machine learning methods are available, in this thesis clustering and kNN method are proposed to perform analytics on energy consumption data, firstly to size and define most suitable operation mode for energy system, then to perform short forecasting.

As previously cited, firstly clustering is proposed to perform analytics on energy demand data, dividing dataset into homogenous groups. The main scope is to analyse which cogeneration technology suits better the energy demand avoiding a mismatch between production and demand (and consequently, heat losses). On the other side, an improvement in operation strategy and definition of the energy storage system defining average consumptions curves can be useful. A case study of an industrial facility where both electricity and heat are required is used to validate the method proposed: it suggests not only the most suitable cogeneration technology for the observed energy demand but also heat storage and improvements on operation strategy to avoid heat losses and, consequently, increasing energy efficiency. It is demonstrated that primary energy saving between 2-6% could be achieved.

In the final part of the thesis, clustering and kNN method are applied to analyse data and perform short forecasting on energy consumption. These methods have been already used to perform forecasting when it is based mainly on historical observations and it is not easy to define mathematical models such as in the case of an industrial process. Novelty on data normalisation and hyperparameters definitions are proposed to decrease error on forecasting. The proposed method is then applied to a case study to verify its performances: it was able to forecast electricity consumption with a percentage error lower than 5% whereas an estimated error of 13% can be estimated if novelties on data normalisation are not applied.

SUMMARY

LIST OF ABBREVIATIONS

AEL	Alkaline Electrolysis
AFC	Alkaline Fuel Cell
AI	Artificial Intelligence
ANN	Artificial Neural Network
ARIMA	Autoregressive Integrated Moving Average
ASHP	Air Source Heat Pumps
CHP	Combine Heat and Power
CPO	Catalytic Partial Oxydizer
DNA	Dynamic Network Analyser
FC	Fuel Cell
GIS	Geographic Information System
GSHP	Ground Source Heat Pumps
HHV	High Heating Value
H/P	Heat to Power Ratio
kNN	k-Nearest Neighbors
ICE	Internal Combustion Engines
IT	Information Technology
IIOT	Industrial Internet Of Things
MCFC	Molten Carbonate Fuel Cell
PAFC	Phosphorous Acid Fuel Cell
PCA	Principal Component Analysis
PE	Primary Energy
PEFC	Proton Exchange Fuel Cell
PEM / PEMEL	Proton Exchange Membrane Electrolysis
PLC	Programmable Logic Controller
RES	Renewable Energy System
RSOC	Reversible Solid Oxide Cell
SD	Standard Deviation
SOEC	Solid Oxide Electrolysis Cells
SOEL	Solid Oxide Electrolysis
SOFC	Solid Oxide Fuel Cell
SQL	Structured Query Language
SVM	Support Vector Machine

LIST OF FIGURES

LIST OF TABLES

1 FUEL CELL FOR ENERGY GENERATION

1.1 ENERGY GENERATION IN INDUSTRIAL FACILITIES

Industry is one of the sectors with the highest consumption of energy: according to Figure 1, about 25 % of the final energy consumption in European Union is related to industry. Moreover, industry may have different types of energy request: meanwhile electricity is used mainly by electric motors and heaters, in some cases cooling could be required (especially by industries related to food processing), or heat at different temperature (hot water or steam). One of the more effective ways to increase energy efficiency and to save money is to implement cogeneration (or polygeneration) systems: primary energy consumption can decrease if two or three types of energy (for example electricity and heat) are produced simultaneously by one single plant instead of separate production.

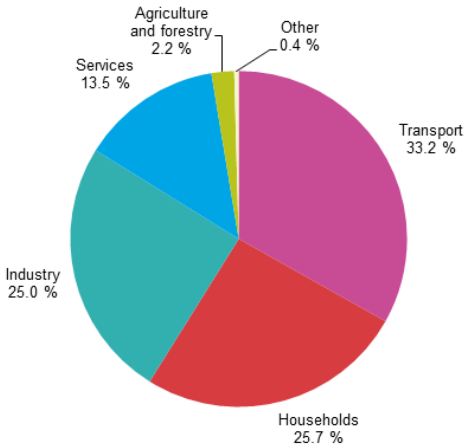


Figure 1. Final energy consumption by sector, EU-28, 2016 (% of total, based on tonnes of oil equivalent) [1]

Figure 2 displays the average H/P based on energy demand by industrial sector ([2]) compared to H/P of the traditional cogeneration technologies ([3]): as a matter of fact, there is often a mismatch between energy production and consumption, so integration systems (such as boiler, steam generator and/or electricity from grid) are required. It is possible to appreciate also that fuel cells have the lowest H/P range (0.5-1) with only one industrial sector (*Printing and publishing*) inside this range. Internal combustion engine has, instead, more industrial sectors (*Apparel and other textile products, Lumber and wood products, Industrial machinery and equipment, Instruments and related products and Printing and publishing*) inside its H/P range (0.83 - 2).

In this part of the thesis, fuel cells (firstly considering solid oxide fuel cells (SOFC), then analysing also reversible solid oxide cells) are analysed for industrial purpose because of the following positive characteristics: the highest efficiency on energy conversion (both on electricity only and overall cogeneration) and their modularity. Table 2 based on reference [3] compares

cogeneration technologies on electric and overall efficiency (based on HHV), typical capacity, H/P range, partial load and type of fuel used. Fuel cell could reach 60 % on electricity efficiency meanwhile internal combustion engines (ICE) and steam turbines could reach 40 %. Overall efficiency reaches 80 % on HHV. In the next chapters, firstly fuel cells (and SOFC in particular) are presented with a detailed description of the model used for simulation. An improvement of the model is presented with an innovative heat recovery system based on air source heat pump, in order to increase the overall efficiency on cogeneration. Successively, an energy conversion system based on SOFC and reversible solid oxide cells (RSOC) is proposed to adjust H/P ratio of energy request meanwhile hydrogen could be produced. A case study of a paper mill is considered where efficiency of energy generation system is increased using SOFC – RSOC system.

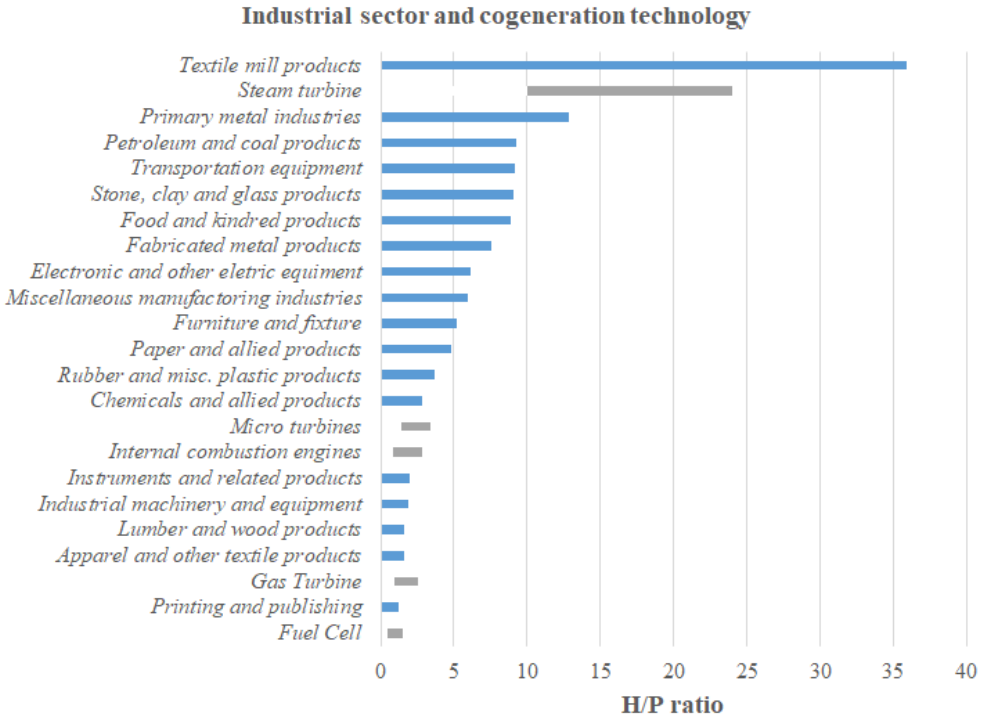


Figure 2. H/P ratio of average energy demand by industrial sector ([2]) and typical cogeneration technologies ([3])

	ICE	**Steam Turbine**	**Gas Turbine**	**Micro Turbine**	**Fuel Cell**
Electric Efficiency (HHV)	27-41%	5-40%	24-36%	22-28%	30-63%
Overall efficiency (HHV)	77-80%	Near 80%	66-71%	63-70%	55-80%

	ICE	Steam Turbine	Gas Turbine	Micro Turbine	Fuel Cell
Typical capacity (MWe)	0.005-10	0.5 – several hundred MW	0.3-300	0.03-1	0.001-2.8 commercial use
H/P range	0.83-2	10-14.3	0.9-1.6	1.43-2	0.5-1
Partial load	Yes	Yes	Poor	Yes	Possible
Fuels	Natural gas, biogas, LPG, sour gas, industrial waste gas, manufactured gas	All	Natural gas, synthetic gas, landfill gas and fuel oils	Natural gas, sour gas and liquid fuel	Hydrogen, natural gas, propane and methanol

Table 1. Overview of cogeneration technologies [3]

1.2 FUEL CELLS

1.2.1 GENERAL INTRODUCTION TO FUEL CELL

Fuel cells (FC) are power devices that convert chemical energy inside a fuel directly in electricity with a high efficiency [4]. Hydrogen or hydrocarbons as fuel (depending on the type of fuel cells) and an oxidant (air) are used. Basically, a fuel cell's generator is made by these elements:

- Unit cell, where the electrochemical reactions take place;
- Stacks, where many units are electrically connected to obtain the desire power output;
- Auxiliaries such as air compressors, pumps, fuel processor (some kinds of fuel cells need it) and a DC/AC converter.

The core of the entire system is the unit cell (Figure 3) where fuel and oxidant react and the products are electricity (DC), heat and some gases with minimal pollutant. It is divided by:

- Anode, the part of the cell fed with the fuel;
- Cathode, the part of the cell fed with the oxidant (usually air);
- Electrolyte, who physically separate anode and cathode allowing the transportation of ions.

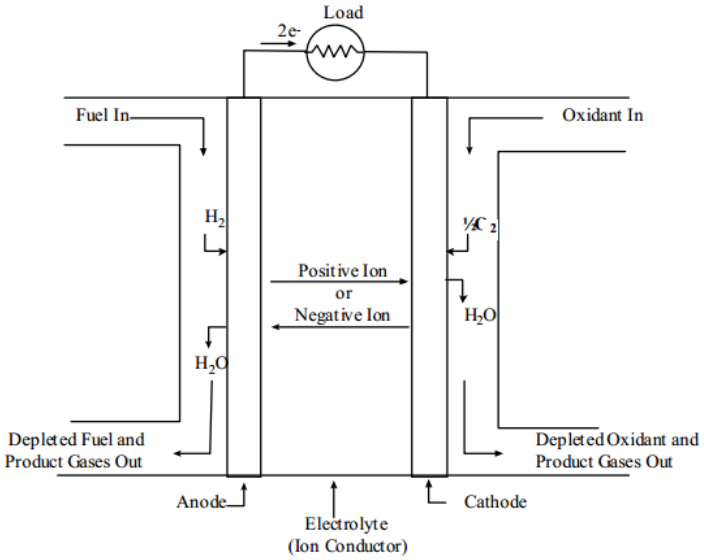


Figure 3. Representation of a Unit Cell [4]

Anode and cathode could be made with a catalytic material such as Platinum to allow a higher rate of reaction; some types of fuel cells work at low temperature, that is at low rate. The type of electrolyte of the unit cell is important because it defines:

- range of working temperature of the unit cell (it could vary from 40-80 °C of polymeric membranes to over 600 °C of ceramics);
- type and purity of the fuel (pure hydrogen with a low rate of CO or hydrocarbons);
- type of oxidant (air or air with a low rate of CO_2);
- sensitivity on pollutant (such as CO, sulphurs, ammonia).

A single unit has low power (usually about 15-20 W), so in order to reach the requested power many cells are connected electrically in series and parallel.

1.2.2 DESCRIPTION OF DIFFERENT TYPES OF FUEL CELLS

The first fuel cell was developed by *Sir William Robert Groove* (1811-1896): it had an electrode made of platinum put into nitric acid and another one made of zinc into copper sulphate, generating a current of 12 A at 1.8 V. Until now, different kinds of fuel cell (FC) were improved especially for aerospace: for example, a type of alkaline FC was created by *Francis Thomas Bacon* and used for Apollo Program of NASA, or proton exchange fuel cell (PEFC) was improved by GE and used for Gemini Program.

The modern fuel cells are divided by the material of the electrolyte: proton exchange fuel cell (PEFC), alkaline fuel cell (AFC), phosphoric acid fuel cell (PAFC), molten carbonate fuel cell (MCFC), and solid oxide fuel cell (SOFC). This allows using different fuels and defining different operating temperatures. Table 2 makes a summary of the different FC developed.

Table 2. Characteristic of different types of FC [4]

	PEFC	**AFC**	**PAFC**	**MCFC**	**SOFC**
Electrolyte	Hydrated polymeric ion exchange membranes	Potassium hydroxide in asbestos matrix	Immobilized liquid phosphoric acid in SiC	Immobilized liquid molten carbonate in $LiAlO_2$	Perovskites, Y_2O_3-stabilised ZrO_2 (YSZ) (ceramics)
Operating Temperature (°C)	40-80 °C	65-220 °C	205 °C	605 °C	600-1000 °C
Stack elect. efficiency (% of LHV)	32%-40%	36%-45%	43%-50%	43%-55%	50%-60% (SOFC+GT 70%)
Gases at anode	Only H_2 <10 ppm CO	Only H_2 <0.5% CO <50 ppm H_2S	Only H_2 <0.5% CO <50 ppm H_2S	H_2, CO <0.5 ppm H_2S	H_2, CO <0.5 ppm H_2S
External reformer for hydrocarbon fuel	Yes	Yes	Yes	No, for some fuels	No, for some fuels and cell design
External shift conversion of CO to hydrogen	Yes with purification to remove CO	Yes with purification to remove CO	Yes	No	No

	PEFC	AFC	PAFC	MCFC	SOFC
Type of fuel	Natural Gas LPG Methanol Ethanol Gasoline Diesel	The same of PEFC	The same of PEFC plus landfill gas	The same of PEFC plus coal gas	The same of PEFC plus coal gas
Advantages	Rapid start-up, absence of corrosive material, high current densities (2 W/cm^2)	Excellent performances	Low operating temperature allow to use common construction material, waste heat could be used for cogeneration or bottom cycle	Hydrocarbons and CO could be used as fuel and no expensive electro-catalysts are needed, the high temperature of waste heat allow bottom cycle to improve overall efficiency	Hydrocarbons could be used as fuel, high efficiency, high temperature of waste heat allow bottom cycles and cogeneration, cheaper material could be used for cell hardware
Disadvantages	Require water management, quite sensitive to poisoning (CO, sulphurs, ammonia)	Require removing CO and CO_2 from H_2 and also from the air in the cathode	Phosphoric acid is highly corrosive, require expensive material in the stack	Very high corrosive electrolyte require expensive material for cell hardware, low power densities (0,2 W/cm^2)	High temperature cause thermo mechanical stress, thermal mismatches among materials and the thermal cycling.

1.3 SOFC

1.3.1 INTRODUCTION TO SOFC

One type of fuel cells presented in Table 2 is Solid Oxide Fuel Cell (SOFC): electrolyte is a solid, non-porous material usually based on Y_2O_3-stabilised ZrO_2 (YSZ), the anode is a Ni-ZrO_2 and the cathode is Sr-doped $LaMnO_3$. This allows using a high operation temperature (600-1000 °C) and consequently:

- high electrical efficiency;
- high temperature of the waste heat that allow bottom cycles (gas turbine, steam turbine or also Stirling engine) or cogeneration;
- internal reforming, not only hydrogen but also hydrocarbons can be used;
- lower sensibility on pollutant, differently from other fuel cells (like as PEFC and AFC), CO could be used as fuel and it is not a pollutant.

The main disadvantage is related to the high operation temperature, as problems can occur due to the mismatching between materials with different thermal expansion coefficients, mainly the interconnections and the seal materials. Steady-state working is better than continuous start-stop because it avoids thermal cycles and increases the life time. The research in this field is directed to develop:

- materials with a better electrical and thermo-mechanical performances at 1000 °C;
- materials (such as mixture of ceramics and metals) with lower operating temperature, but with higher power density and efficiency.

The power range is from 1 kW to 10 MW, so SOFC can be used for household application, auxiliary power unit and as power plant with high efficiency.

The most diffused cells developed today have tubular and planar form. The first type, made by Siemens-Westinghouse, is the best-known and a cross section is represented in Figure 4. The cathode is produced by extrusion and sintering, then the electrolyte is applied by electrochemical vapour deposition (EVD). The anode is produced by metallic Ni and YSZ by sintering metal particles with a porosity of 20%-40% to allow the mass flow of reactant and product gases; it requires a thermal expansion comparable to the other cell materials. The cell interconnection is made by doped lanthanum chromite, and it must be chemical unreactive to resist to both anode and cathode gases. All the cell components must:

- be capable of withstanding thermal cycling;
- not be affected on electronic conductivity by interdiffusion of ionic species;
- permit only electronic conduction and interdiffusion of ionic species at 1000 °C.

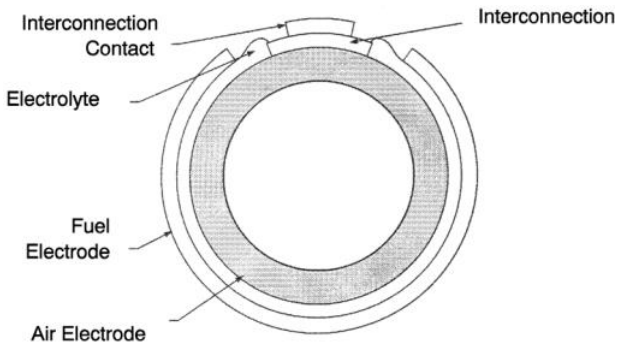


Figure 4. Cross-section of cylindrical Siemens-Westinghouse SOFC [4]

Planar SOFCs are based on planar cells (Figure 5). Usually the shape of the cell is rectangular or circular: the gases flow in counter-flow or cross-flow in the first case, or in co-flow in both of them. Moreover, it is possible to distinguish the different cells according to structural support. The early cells were electrolyte-supported, so they required a thick electrolyte around 200 μm with both electrodes at 50 μm. The anode-supported cells are possible thanks to advances in manufacturing techniques: the anode is 0.5-1 mm thin, the electrolyte is around 3 to 15 μm and the cathode is 50 μm (it gives also the difference in thermal expansion between anode and electrolyte). This type of fuel cell has very high power density, from 1.8 W/cm^2 in laboratory tests to 600-800 mW/cm^2 in commercial conditions. A cathode-supported fuel cell would also be possible, but the mass flow limitation and the manufacturing challenges allow lower performance than the anode-supported. A kind of metal interconnection-supported fuel cell has been studied to minimize the use of expensive ceramic materials and mass flow resistance. The problems are finding a materials' combination and a manufacturing process that avoid corrosion and deformation of the metal and interfacial reactions during manufacturing and operation. Three different materials are used as interconnection:

- lanthanum or yttrium chromite (ceramic); they work pretty well at high operating temperature (900 °C - 1000 °C), they are chemically stable with thermal expansion. Unfortunately, they are costly and mechanically weak;
- Cr-based or Ni-based superalloys for intermediate-high operation temperature (800 °C - 900 °C). They are chemically stable, but they require a coating to prevent Cr-poisoning. Moreover, this technology features high cost due to the difficulty to form Ni and Cr;
- ferritic steel for intermediate operation temperature. The material is cheaper than the other ones, but it needs a coating to improve corrosion resistance and conductivity during thermal cycles because it is chemically unstable.

The power density of planar SOFC (about 300-500 mW/cm^2) allows to engineer small-scale power and APU application, with the possibility to reach a customization for high-volume application.

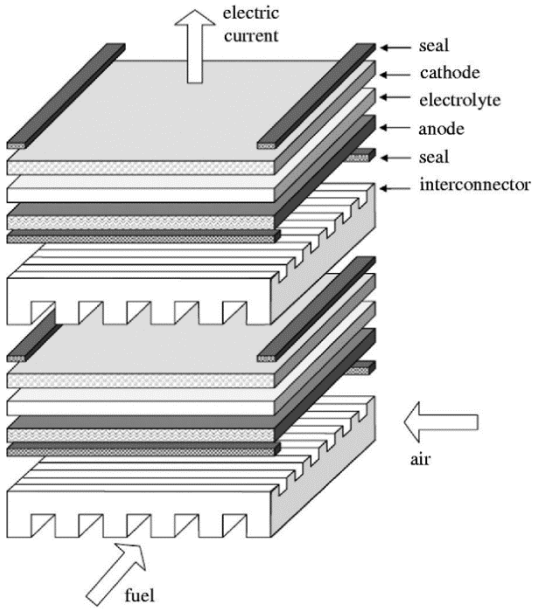


Figure 5. Representation of a planar fuel cell with a rectangular shape [5]

1.3.2 THERMODYNAMIC OF SOFC

In a fuel cell an electrochemical reaction occurs, fuel is oxidized with the production of electricity and exhausted gases that could be used for heat recovery (Table 3).

Anode Reaction	Cathode Reaction
$H_2+O^{2-}\rightarrow H_2O+2e^-$ $CO+O^{2-}\rightarrow CO_2+2e^-$ $CH_4+4O^{2-}\rightarrow 2H_2O+ CO_2 +8e^-$	$½\ O_2+2e^-\rightarrow O^{2-}$

Table 3. Chemical reaction in a SOFC [4]

When it is operative, at the anode the oxidation of H_2 and CO takes place (with the loss of electrons), and at the cathode the reduction of the O_2 occurs. The current I of the electric circuit depends on the flow of electrons. The voltage of a fuel cell could be defined by Nernst's Equation. In ideal condition and with I=0, the ideal voltage at the working temperature T is (Eq. 1):

$$V_{id,T} = \frac{RT}{nF} ln \frac{\prod(r)}{\prod(p)} \qquad \text{Eq. 1}$$

where V_{id} is the voltage at standard ambient temperature and pressure (25 °C and 101325 Pa), R is ideal gas constant, n is the electrons flow, F is the Faraday constant, $\Pi(r)$ and $\Pi(p)$ are the product of reactants and product of fugacity respectively. This ideal value won't be reached even at $I=0$ due to irreversibility losses. Three main phenomena cause a reduced efficiency of fuel cell with respect to the ideal value (Figure 6):

- activation-related losses, related to chemical changes on electrodes surface or gas absorption. They could be described by Tafel equation where *b* is called *Tafel gradient* and the unit of measurement of *i* is mA/cm^2 (Eq. 2).

$$\beta_1 = a + b * log(i) \qquad \text{Eq. 2}$$

- ohmic losses, related to electric resistivity of electrolyte and of electrodes. They could be described using Ohm law (Eq. 3).

$$\beta_2 = R_{int} * I \qquad \text{Eq. 3}$$

- mass-transport-related losses, related to operation condition. When the current density is too high, the reactant diffusion into the electrolyte and the replacement of the products become too slow, so a concentration gradient with a loss of voltage happens. They could be described by Eq. 4 where i_L is the limitation current, the highest speed at which it is possible to feed reactants to the fuel cell.

$$\beta_3 = \frac{RT}{nF} ln(1 - \frac{i}{i_L}) \qquad \text{Eq. 4}$$

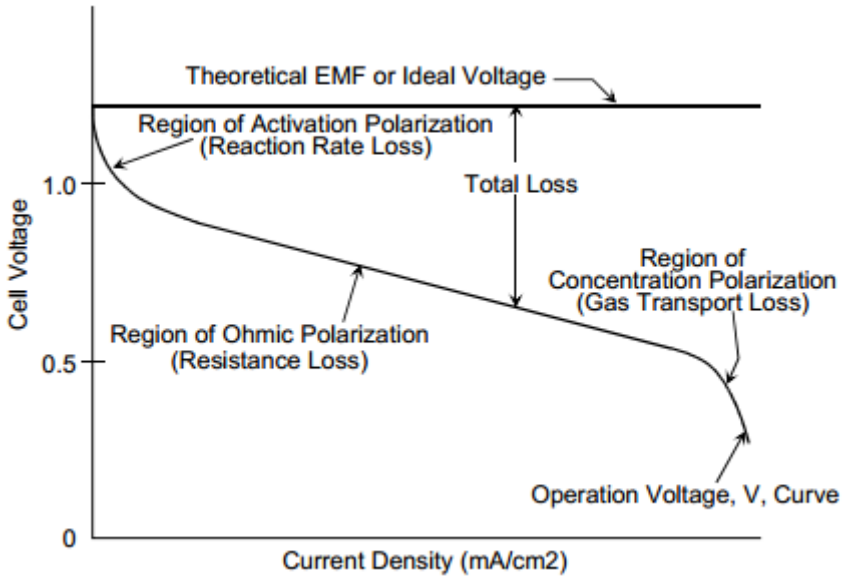


Figure 6. Voltage loss related to current density [4]

1.3.3 SOFC SYSTEM SIMULATION

A SOFC system is simulated using a tool developed for thermodynamic simulation at Technical University of Denmark (DTU), DNA – Dynamic Network Analyser [6]. It is developed in Fortran language and by using this tool it is possible to define thermodynamic systems based on different

components (such as gasifier, steam turbines, gas turbines, burner, fuel cell, absorption chiller etc), and it allows to perform both steady-state and dynamic simulations.
A SOFC model was developed by prof. Marvin Mikael Rokni (formerly Masoud Rokni). It was firstly presented in [7], and then it was proposed with different fuels such as ethanol [8]–[10], methanol [8]–[10], ammonia [8]–[10], DME [8]–[10], woodchips gasification [11], biomass [12] or municipal waste gasification [13]. As previously mentioned, SOFC could be coupled with a bottom cycle to increase electricity production, for example by using Stirling engine [8], [10], [12]–[14], steam turbine [7] or steam injected turbine [11]. Even if previous references present a SOFC system for electricity production with a net power in the order of megawatt, SOFC was studied also for small power application with ground source heat pumps (GSHP) in [10], [15]–[17] thanks to its modularity. In this thesis, a modular SOFC system using natural gas as fuel ([10], [18], [19]) is proposed and simulated: it is based on SOFC stacks, a catalytic partial oxidizer (CPO), heat exchangers to pre-heat fuel and oxidant, air compressors, a desulfurizer and a burner (Figure 7).

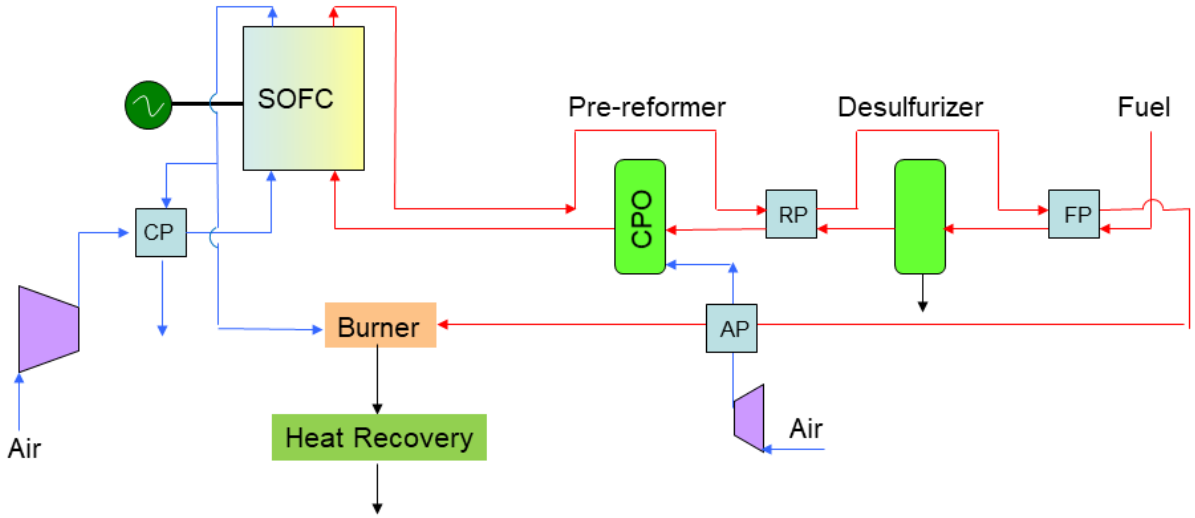


Figure 7. SOFC system representation with heat recovery

The desulfurizer is necessary to avoid sulfur poisoning of the SOFC stack, CPO cracks hydrocarbons into CO and H_2, air compressors compress air for cathode and CPO. CP is the cathode pre-heater: it uses exhausted gases to pre-heat inlet air for the SOFC. Fuel is pre-heated into FP and RP heat exchanger that use exhausted gases from the anode. Electricity is converted by an AC/DC converter with an efficiency assumed to be 92 %, meanwhile auxiliaries' consumptions are assumed to be 1.5 % of the total electricity production. Exhausted gases are burned to convert the unused fuel into heat that could be recovered to produce steam. The latter could feed bottom cycles or could be used for space heating. In the next paragraph (1.3.4), the mathematical model of SOFC stack is described.

1.3.4 SOFC STACK DEFINITION

SOFC stacks are simulated using a zero-dimensional model ([12]) which describes power using experimental data. Gases temperature is considered varying between 650 °C and 800 °C. Once defined Nernst ideal reversible voltage E_{Nernst}, activation polarization losses ΔE_{act}, ohmic losses ΔE_{ohm} and concentration losses ΔE_{conc}, the operational voltage of the cell E_{cell} can be defined as (Eq. 5):

$$E_{cell} = E_{Nernst} - (\Delta E_{act} + \Delta E_{ohm} + \Delta E_{conc}) \quad \text{Eq. 5}$$

Under the assumption that only H_2 is used, E_{Nernst} could be defined as (Eq. 6):

$$E_{Nernst} = \frac{-\Delta g_f^0}{n_e F} + \frac{RT}{n_e F} ln(\frac{p_{H_2,tot}\sqrt{p_{O_2}}}{p_{H_2O}}) \quad \text{Eq. 6}$$

$$p_{H_2,tot} = p_{H_2} + p_{CO} + 4p_{CH_4} \quad \text{Eq. 7}$$

Where Δg_f^0 is the Gibbs free energy for H_2 reaction at standard temperature and pressure, $p_{H_2,tot}$, p_{O_2} and p_{H_2O} are the partial pressure for hydrogen, oxygen and water respectively (Eq. 7), F is the Faraday constant and R the universal gas constant.

Activation polarization is described using Butler-Volmer equation (Eq. 8):

$$\Delta E_{act} = \frac{RT}{(0.001698T - 1.254) * F} \sinh^{-1}(\frac{i_d}{2 * (13.087 * T - 1.096 * 10^4)}) \quad \text{Eq. 8}$$

Where T and i_d are respectively the operating temperature and current density. Ohmic losses (Eq. 9) can be defined as related to electrical conductivity of electrodes (σ_{an} for anode, σ_{ca} for cathode), ionic conductivity of the electrolyte (σ_{el}), thickness of these elements (t_{an} for anode, t_{ca} for cathode and t_{el} for electrolyte) and current density (i_d):

$$\Delta E_{ohm} = (\frac{t_{an}}{\sigma_{an}} + \frac{t_{ca}}{\sigma_{ca}} + \frac{t_{el}}{\sigma_{el}}) i_d \quad \text{Eq. 9}$$

Thicknesses are assumed to be t_{an} = 600 µm for the anode, t_{el} = 50 µm for the electrolyte and t_{ca} = 10 µm for the cathode. Conductivity is assumed to be $\sigma_{an} = 10^5$ for the anode, meanwhile conductivity of the cathode and of the electrolyte are related to the operating temperature T and are defined by Eq. 10 and Eq. 11 respectively:

$$\sigma_{ca} = \frac{5.76 * 10^7}{T} \exp(-\frac{0.117}{8.617 * 10^{-5} T}) \quad \text{Eq. 10}$$

$$\sigma_{el} = 8.588 * 10^{-8} * T^3 - 1.101 * 10^{-4} * T^2 + 0.04679 * T - 6.54 \qquad \text{Eq. 11}$$

Concentration losses (ΔE_{conc}) occur when insufficient amounts of reactants are transported to the electrodes, mainly at high current densities for anode-supported SOFC. Experimental data were used to validate the model introducing anode limiting current, a parameter based on anode porosity and tortuosity. Equation proposed is (Eq. 12):

$$\Delta E_{conc} = B\left(\ln\left(1 + \frac{p_{H_2} * i_d}{p_{H_2O} * i_{as}}\right) - \ln\left(1 - \frac{i_d}{i_{as}}\right)\right) \qquad \text{Eq. 12}$$

where *B* is the diffusion coefficient defined using a calibration technique (Eq. 13):

$$B = \left(0.008039 * X_{H_2}^{-1} - 0.007272\right) * \frac{T}{T_{ref}} \qquad \text{Eq. 13}$$

where X_{H_2} is the mass reaction rate of H_2 and T_{ref} is reference temperature (1023 K). Anode limiting current is defined as (Eq. 14):

$$i_{as} = \frac{2 * F * p_{H_2} * D_{bin} * V_{an}}{R * T * t_{an} * \tau_{an}} \qquad \text{Eq. 14}$$

where V_{an} is the anode porosity (assumed 30%), τ_{an} is the anode tortuosity (2.5) and D_{bin} the binary diffusion coefficient. D_{bin} is also calibrated using experimental data (Eq. 15):

$$D_{bin} = \left(-4.107 * 10^{-5} * X_{H_2} + 8.704 * 10^{-5}\right) * \left(\frac{T}{T_{ref}}\right)^{1.75} * \frac{p_{ref}}{p} \qquad \text{Eq. 15}$$

where p_{ref} is the reference pressure (1.013 bar). Current density i_d is defined using Faraday law (Eq. 16):

$$i_d = \frac{\dot{n}_{H_2} * 2 * F}{A} \qquad \text{Eq. 16}$$

A is the physical area of the cell (144 cm^2). Each stack is composed by 70 cells, each stack has a power of 1 kW, and the number of stacks varies with the nominal power of the SOFC to be achieved. Figure 8 represents SOFC voltage over current density, comparing simulation data with experimental data ([12]): it is possible to appreciate that the proposed model describes perfectly experimental data.

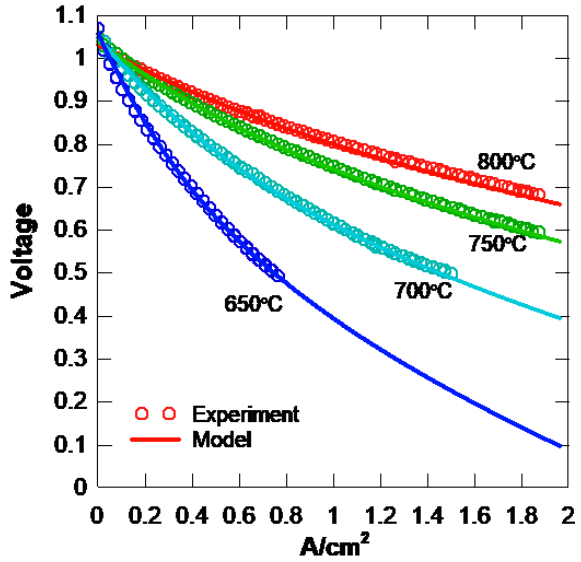


Figure 8. SOFC voltage over current density, model versus experimental data ([12])

1.4 SOFC – Air Source Heat pump (ASHP) system for advanced heat recovery

1.4.1 Introduction to the system

As previously mentioned ([7], [11]), SOFC system can be used to produce steam that is a valuable thermal vector in industrial facilities (for example it is used into paper mills to dry paper). Figure 8 illustrates the SOFC system proposed in [7] where exhausted gases are used to produce steam for a Rankine cycle in three different stages: economizer, evaporator and super-heater.

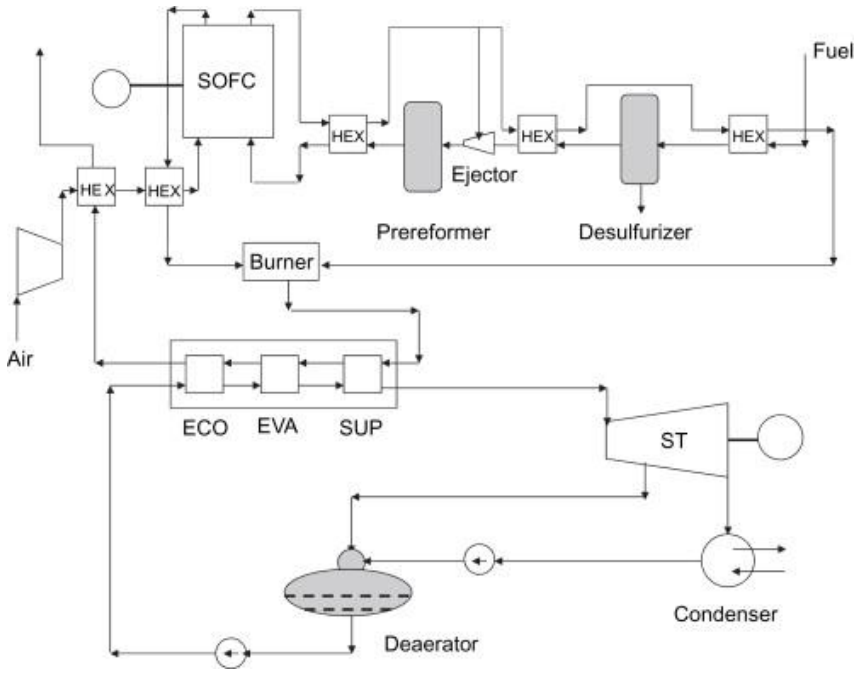


Figure 9. Example of SOFC system with steam production for a Rankine cycle [7]

On the other hand, exhausted gases after heat recovery still have an enthalpy content that is unused. Industrial processes do not require heat at only medium-high temperature: for example, processes like fermentation uses heat at temperature of 50 °C (Figure 10). Heat pumps have been already proposed to increase energy efficiency on heat production for these processes: reference [20] analyses how heat pumps can be used to produce thermal energy for industrial processes at different temperatures.

Industrial waste heat could be used not only in industrial processes, but also for district heating: in reference [21], a tool to analyse the possibility of using waste heat for district heating varying cost of fuel, electricity and distance for the transfer to district heating is developed. In [22], [23] it is analysed the possibility of heat recovery of an industrial area in the north of China. In [24] it is studied the technical, economic, institutional and environmental feasibilities of using low-level residual industrial waste heat for the district heating of Delft (The Netherlands). In [25], a district heating system using waste heat from industries, waste incineration and a cogeneration plant is proposed and analysed to improve energy saving and to achieve economic saving.

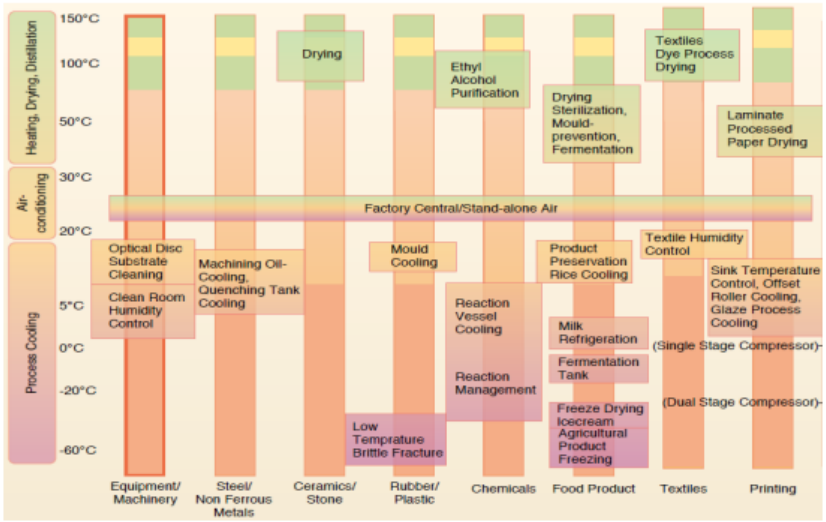


Figure 10. Industrial processes thermal request divided by temperature ([20])

In this study, air source heat pump (ASHP) combined with SOFC is proposed. The main idea is to increase heat production, and consequently the coefficient of performance (COP) of the heat pump: exhaust gases exiting the fuel cell are firstly used to satisfy medium - high temperature heat request (such as steam production), then they are mixed with external air using an adiabatic mixer to feed the evaporator of the heat pump with the aim of increasing energy efficiency of the latter (Figure 11).

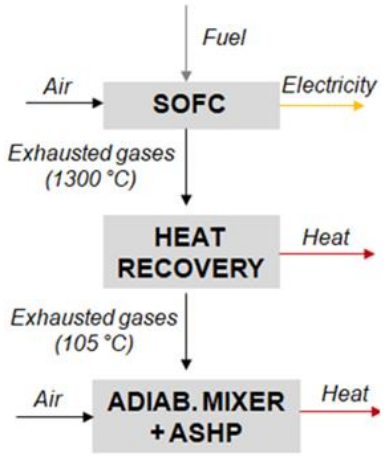


Figure 11. SOFC + ASHP system scheme

ASHP are cheaper compared to other types of heat pumps such as ground source heat pumps, but efficiency decreases if freezing of evaporator (which is one of the drawbacks for air source heat pump in area like North Europe) occurs. System configuration using air and exhausted gases of SOFC decreases the possibility of freezing that could occurs when temperature is around 0 °C and relative humidity is sufficiently higher than 50 %. A modular system based on a SOFC system of 50 kW of nominal power and 7 kW of ASHP heating power is proposed. Simulations are performed varying external air temperature, air humidity and SOFC nominal power.

1.4.2 SOFC SYSTEM

As previously mentioned, the system is proposed to recover heat from waste gases of SOFC. A modular system of 50 kW SOFC nominal power (50 stacks of 1 kW each) and 7.7 kW ASHP heating power is proposed. SOFC system is based on the scheme analysed in section 1.3.3, natural gas is proposed as fuel even if ammonia, DME, ethanol or methanol could be used just changing fuel processor components ([8], [9]). When natural gas is used, desulphurizer and CPO are necessary: the first is required to prevent sulphur poisoning of the stack, the second to crack hydrocarbons. Figure 12 represents a detailed scheme of the SOFC system and interconnection with the ASHP. An adiabatic mixer is used in order to mix exhausted gases and inlet air. Simulation of SOFC at full nominal power (50 kW) reports an electric efficiency of 53 % and a thermal efficiency 42.86 %, meanwhile waste gases after heat recovery are at 105 °C. Table 4 resumes the main results of the thermodynamic benchmarks of SOFC system simulation at full load.

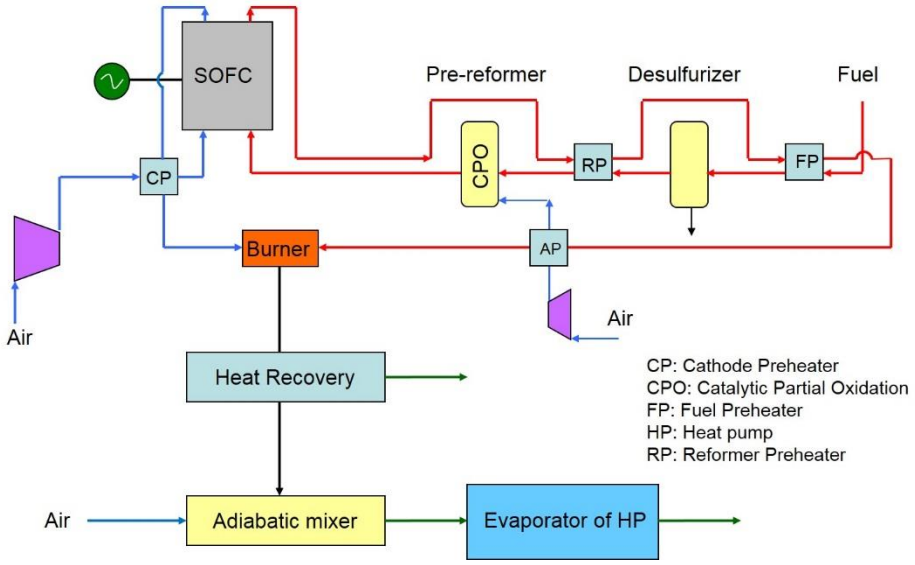


Figure 12. Representation of SOFC system. The air mixer to partially recover heat from the exhausted gases of the SOFC is connected after the Heat Recovery by point 1 (see previous Figure 15).

Parameter	Value
SOFC (50 kW, full load), thermal efficiency	$\eta_{thermal,SOFC} = 0.4286$
SOFC (50 kW, full load), electrical efficiency	$\eta_{electrical,SOFC} = 0.5299$
SOFC (50 kW, full load), heat to power ratio	H/P = 0.8088
SOFC auxiliaries consumption, efficiency on electrical output	$\eta_{trans} = 0.9068$
Exhausted gas temperature	105 °C

Table 4. Efficiencies of the different components of the SOFC system.

Desulphurizer not only prevents stack poisoning, but also sulphur inside exhausted gases: Table 5 shows the results of exhausted gas analysis. Note that sulphur formation in any form (such as SO_x) is avoided and, consequently, if gas mixture is condensed, condensate may not include any acid: ASHP does not require heat exchanger in acid resistant material such as stainless steel to prevent corrosion by acid condensate. At full-load, it is calculated that the humidity ratio (water mass versus dry air mass) is 0.342 kg_{water}/kg_{dry_air}.

Composition	Gas percentage		Composition	Gas percentage
N_2	56.76%		Ar	0.59%
Water	24.37%		NO	0.00%
CO_2	12.90%		SO_2	0.00%
O_2	5.37%		NO_2	0.00%

Table 5. Mass composition of SOFC exhausted gas at full load, gas temperature 105 °C.

1.4.3 ASHP

ASHP plays an important role in the system: it recovers heat from exhausted gases at (relative) low temperature. ASHPs have variable COP during the year, typically lower during colder periods when freezing of evaporation section may occur. It is advisable to consider that in climates where conditions with air temperature just above 0 °C (especially between 5 and 7 °C depending on the design of the finned coil) and relative humidity above 50 % are more frequent, possibility of freezing of the outdoor heat exchanger (evaporator) may lead to a decrease in seasonal performance of the heat pump (ice has poor heat transfer capability and reduces available area for air and so air mass flow rate). For such a reason defrost of evaporator section is periodically necessary. Defrosting can be performed by an auxiliary heat source (electrical resistance or gas burner) or reversing the cycle. In any case, defrosting is quite penalizing for the heat pump energy performance, as it increases its energy request. The authors propose to mix the exhausted gases exiting the SOFC (Figure 12 and point 1 in Figure 15) with outdoor air (point 2 in Figure 15) with the aim to enhance temperature of inlet air at the ASHP evaporator (point 3 in Figure 15) in order to prevent ice formation.

Heat pumps could be simulated using technical norms, for example UNI 11300-4 to consider different working temperature at condenser/evaporator and using EN 14825 for partial load of the heat pump in heating mode ([26]–[28]). In this study, it is proposed to simulate ASHP using a regression of technical datasheet (Figure 13, Figure 14, and Table 6) from [29].

As previously mentioned, when external air relative humidity is higher than 50 % and temperature is just above 0 °C, freezing of evaporator may occur. Typically, freezing rate is maximum when air temperature is around 7 °C. A frost factor is considered [30] as multiplying penalty factor to decrease COP of the heat pump to take into account the periodic defrosting. Reference proposes a value that is defined as a function of outdoor air temperature (in the range of –10 to 10 °C) and relative humidity (in the range of 50-100 %): the penalty factor is lower given a lower outdoor air temperature (up to values just above 0 °C), and the higher the relative humidity [30].

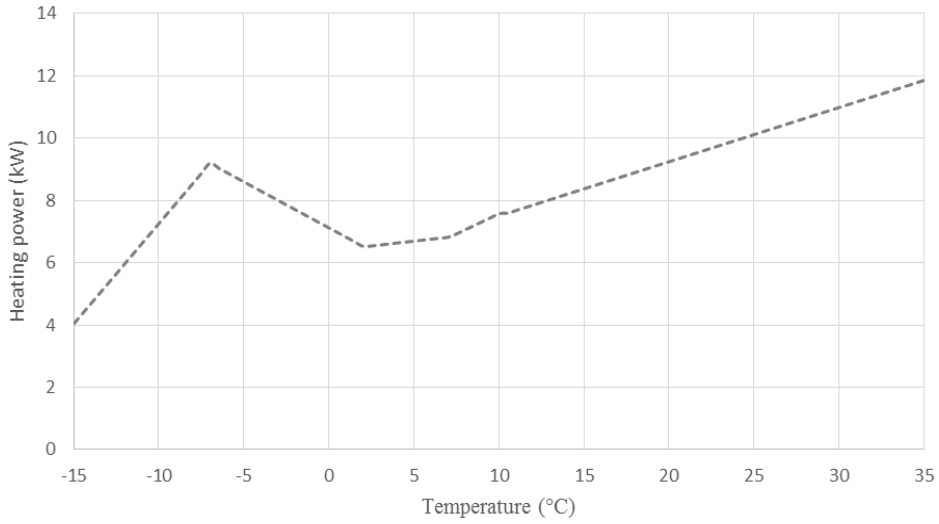


Figure 13. Technical datasheet, relation between nominal heating power and external air temperature ([29]).

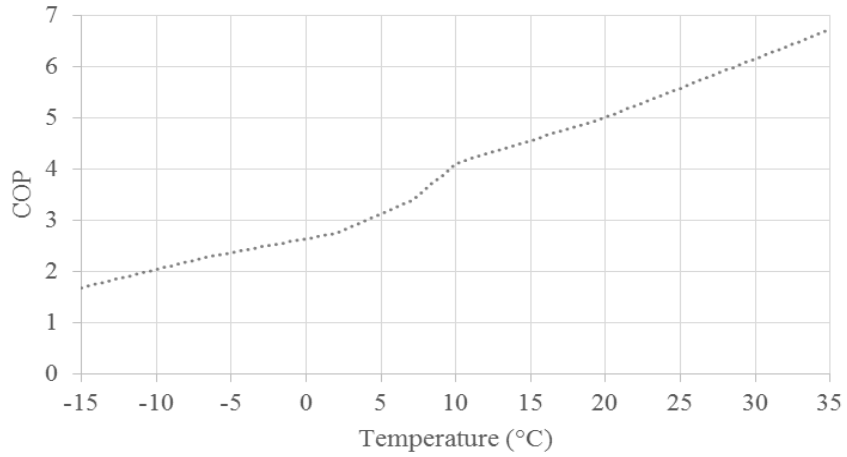


Figure 14. Technical datasheet, relation between nominal COP and external air temperature ([29]).

Parameter	Value
ASHP Nominal Condition	External Air 2 °C - Condenser outlet Water 45 °C (A2W45)
Pnom – Nominal heating power	7.7 kW
COPnom – Nominal COP	2.75

Table 6. Nominal conditions of ASHP ([29]).

1.4.4 ADIABATIC MIXER AND EVAPORATOR

As previous mentioned, the proposal is mixing air with off-gases from the SOFC plant. The aim is to increase air temperature at evaporator inlet as much as possible to avoid freezing of evaporator and increasing the COP. Figure 15 represents the proposed air mixing system: SOFC

exhausted gases (point 1) and outdoor air (point 2) are mixed in an adiabatic mixer. In some cases, humidity condensation may also occur (point 31). Mixture (point 3) is used in the evaporator delivering heat to the ASHP and then is discharged (point 4). Humidity condensation may occur also into the evaporator (point 41). It is supposed that no auxiliary flow inducing system such as a fan is necessary: the head pressure available from the fan of the heat pump model here referred (around 80 Pa) is supposed to be adequate to face the pressure drop of the mixer (and of the ducts and evaporator).

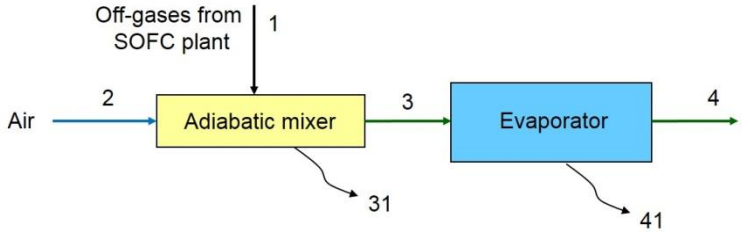


Figure 15. Air mixing system. Curves pointing down represent possible water condensation after the air heat exchange respectively in the mixer (point 31) and the evaporator (point 41).

A mathematical model is proposed to describe the components of the system according to equations regarding wet air proposed in [31]. Firstly, it is described the adiabatic mixer (Eq. 17):

$$\begin{cases} m_{DA,1} + m_{DA,2} = m_{DA,3} \\ m_{DA,1} * W_1 + m_{DA,2} * W_2 = m_{DA,3} * W_3 + m_{l,31} \\ m_{DA,1} * h_1 + m_{DA,2} * h_2 = m_{DA,3} * h_3 + m_{l,31} * h_{l,31} \\ W_3 = \min\left(\frac{m_{DA,1} * W_1 + m_{DA,2} * W_2}{m_{DA,3}}, W_{sat,3}\right) \end{cases} \quad \text{Eq. 17}$$

In this system of equations it is imposed:

- conservation of dry air mass ($m_{DA,1}$, point 1), dry exhausted gases ($m_{DA,2}$, point 2) and dry air mixture ($m_{DA,3}$, point 3);
- conservation of water mass flow rate. W_1, W_2, W_3 are respectively the humidity ratio in points 1, 2 and 3, $m_{l,3}$ is the liquid mass flow rate in case of humidity condensation in the adiabatic mixer;
- conservation of energy. h_1, h_2, h_3 are respectively the specific enthalpy in 1, 2, 3, the values depend on air temperature and humidity ratio. $h_{l,3}$ is the enthalpy of condensate water.

$W_{sat,3}$ is the humidity ratio in state 3 in saturation condition, depending only on air temperature in 3 (total pressure is assumed to be 101325 Pa). If $W_{sat,3}$ is lower than weighted average of humidity ratio in states 1 and 2 condensation occurs.

Then evaporator is defined (Eq. 18):

$$\begin{cases} m_{DA,3} = m_{DA,4} \\ m_{DA,3} * W_3 = m_{DA,4} * W_4 + m_{l,41} \\ m_{DA,3} * h_3 = m_{DA,4} * h_4 + m_{l,41} * h_{l,41} + Q \\ W_4 = \min(W_3, W_{sat,4}) \end{cases}$$ Eq. 18

In this system of equations it is imposed:

- conservation of dry air mixture mass. $m_{DA,3}, m_{DA,4}$ are respectively the dry mass flow rate in states 3 and 4;
- conservation of water mass. W_3, W_4 are the humidity ratio values respectively in 3 and 4, $m_{l,4}$ is the liquid mass flow rate in case of humidity condensation in the evaporator;
- conservation of energy. h_3, h_4 are respectively the specific enthalpy in 3 and 4, they depend on air temperature and humidity ratio. $h_{l,4}$ is the enthalpy of condensate water and Q is the heat absorbed by the refrigerant at the heat pump evaporator.

$W_{sat,4}$ is the humidity ratio in 4 in saturation condition, depending only on air temperature in 4. If $W_{sat,4}$ is lower than humidity saturation in state 3 condensation occurs. The heat pump is expected to increase its performances because of the higher air enthalpy at the evaporator inlet. This is due to both the higher temperature and the higher humidity ratio (both sensible and latent terms contribute to the enhancement of enthalpy due to the adiabatic mixer).

1.4.5 SIMULATION OF THE SYSTEM

A steady-state analysis is performed with the aim of studying the energy performance of the system at different operation conditions. The analysis is performed by varying the dry bulb air temperature from –7.5 °C to 15 °C with a step of 2.5 °C, relative humidity from 25 % to 100 % with a step of 25 %, and SOFC nominal electric power from 20 to 50 kW (step of 10 kW). Only two couples of values, namely outdoor air relative humidity and SOFC nominal electric power, are presented in next section 4 to compare very different situations: 25 % - 20 kW, and 100 % - 50 kW. Varying SOFC nominal electrical power is advisable because it affects the exhausted gases flow rate entering the adiabatic mixer. A 7.7 kW nominal heating power ASHP is chosen to have consistent mass flow rates between heat pump evaporator and SOFC.

The simulations allow to calculate air temperature at the evaporator outlet (point 4 in Figure 15), as well as the COP of the heat pump. The main aim of the proposed system is to increase the COP of the ASHP. As previously mentioned, COP is a function of air temperature and relative humidity. A high value of the latter is useful to improve the COP because of the condensation of latent heat that increases the heat exchange inside the evaporator, under assumption that the finned coil surface temperature is not below 0 °C. Unlike, frost may grow in the fins reducing heat exchange between air and refrigerant. As already cited in section 1.4.3, this may occur more frequently when air temperature in the evaporator inlet is between 5 °C and 9 °C. In such case, a defrost factor has to be considered [30] to take into account the penalization of COP due to periodic defrosting of the evaporator with finned coil (e.g. by reversing the cycle). Analysis of COP variation between a

traditional ASHP (without the adiabatic mixer) and the current system is performed and presented for four very different representative cases by a combination of external air relative humidity and SOFC nominal power (25 % - 20 kW, 25 % - 50 kW, 100 % - 20 kW, 100 % - 50 kW). $COP_{variation}$ benchmark is defined as (Eq. 19):

$$COP_{variation} = \left(\frac{COP_{innov,sys}}{COP_{trad,sys}} - 1 \right) \cdot 100 \qquad \text{Eq. 19}$$

$COP_{variation}$ higher than 0 (zero) means that the current system has a higher COP than the traditional one, and therefore ASHP performs better.

Analysis on primary energy saving (*%PES*) is also proposed, wherein the innovative system is compared with a traditional one with separate production of heat (boiler) and electricity (national grid) in terms of primary energy (*PE*). Considering that the system here proposed has a net available electricity generation E_{ava} (that is the difference between SOFC net electric power and ASHP consumption) and a heat generation H_{ava} (that is the sum of heat cogenerated by SOFC and generated by ASHP), *%PES* benchmark is defined as (Eq. 20):

$$\%PES = \left(1 - \frac{PE_{inno,sys}}{PE_{trad,sys}} \right) \cdot 100 = \left(1 - \frac{\frac{E_{ava}}{\eta_{ele}} + \frac{H_{ava}}{\eta_{boiler}}}{F_{SOFC}} \right) \cdot 100 \qquad \text{Eq. 20}$$

where F_{SOFC} is the fuel (primary energy) consumption of SOFC, η_{ele} is the global electric efficiency from grid (assumed to be 0.435), and η_{boiler} is the efficiency of boiler for heat production (assumed to be 0.9). Such a definition is consistent with that of the primary energy saving of cogeneration systems as referenced in the 2012/27/EU Energy Efficiency Directive [32] and Directive 2004/8/EC on promotion of cogeneration [33].

1.4.6 RESULTS - EVAPORATOR OUTLET AIR TEMPERATURE

Firstly, the difference in evaporator outlet temperature (airside) between an ASHP standalone (only ASHP) and ASHP-SOFC integrated system is outlined (Figure 16). In both cases considered as previously described (case 1: 25 % - 20 kW, and case 2: 100 % - 50 kW), temperature at the evaporator outlet is higher in the ASHP–SOFC system because of the positive effect of high temperature of the exhausted gases from SOFC. The higher the SOFC electric power, the higher the temperature difference between the two systems is found to be. The reason is due to the increasing of airflow rate. For example, at 0 °C external air temperature T_4 is 3 °C higher with ASHP-SOFC system with respect to ASHP only in case 1, whereas at same conditions T_4 is 10 °C higher in case 2.

Even if the increasing of temperature of discharged gases (point 4, Figure 15) is proved, COP may increase or not. If evaporator temperature is between –7.5 °C and 10 °C and relative humidity is

higher than 50 %, then the frost factor shall be considered, which can affect the energy performance of the heat pump.

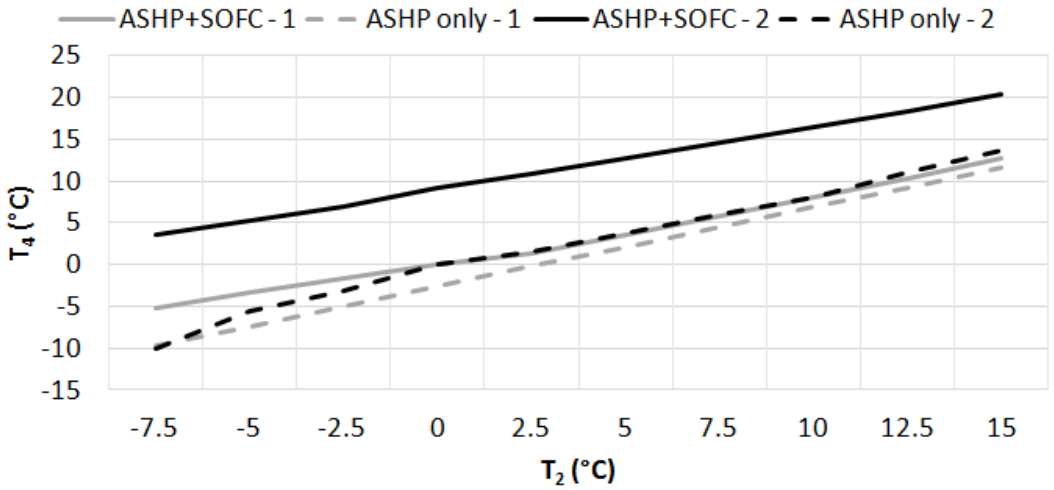


Figure 16. Evaporator outlet air temperature (T_4) in function of external air temperature (T_2) in the two cases (air relative humidity - SOFC nominal electric power). Case 1: 25 % - 20 kW; Case 2: 100 % - 50 kW.

1.4.7 RESULTS - COEFFICIENT OF PERFORMANCE

Figure 17 compares the COP of the presented system with that of an ASHP only, considering the two very different cases previously described: low external air relative humidity with low SOFC nominal electric power (respectively 25 % and 25 kW), and high external air relative humidity with high SOFC nominal electric power (respectively 100 % and 50 kW). It is apparent that the system proposed here is not always advantageous. The latent heat contribution of SOFC exhausted gases may be greater than the sensible one and therefore more frequent defrosting is requested when air temperature is in the critical range (5–9 °C as already stated). The higher weight of the frost factor may decrease the COP of the innovative system. For low humidity and low power (case 1), the proposed system has lower COP than traditional one when external air temperature is lower than about 8.5 °C, ranging from 2 to 3 for ASHP only and from 1.7 to 3 for ASHP-SOFC. However, in the case of 100 % - 50 kW, the higher exhausted mass flow rate (due to the higher electrical power) allows COP of the proposed system to be always higher (between 2.1 and 4.8) than COP of the traditional system (between 1.7 and 4).

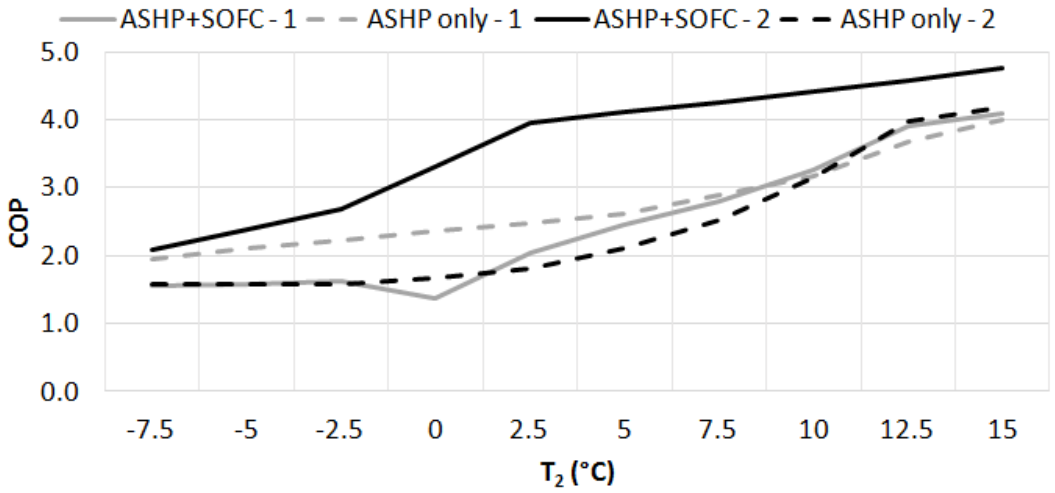


Figure 17. COP in function of external air temperature (T_2) in the two cases (air relative humidity - SOFC nominal electric power). Case 1: 25 % - 20 kW; case 2: 100 % - 50 kW.

Figure 18 depicts $COP_{variation}$ with varying outdoor air temperature for four very different representative cases by a combination of external air relative humidity and SOFC nominal power. $COP_{variation}$ is influenced by:

- inlet air relative humidity. Given a fixed amount of SOFC power, $COP_{variation}$ increases with increasing relative humidity of the air. According to Figure 18, the present system with 20 kW SOFC nominal power has a COP lower than the traditional one ($COP_{variation}$ is lower than 0) at 0 °C and 25 % external air condition. $COP_{variation}$ becomes positive if inlet air has higher relative humidity (e.g. 100 %);
- SOFC electric power. The higher the SOFC power, the higher the mass flow rate of exhausted gases, so the higher the temperature of the gases at the outlet of the adiabatic mixer. This parameter has a strong effect on system performances. Figure 18 shows that for inlet air temperature equal to 2.5 °C $COP_{variation}$ increases from 30 % up to 110 % when considering a SOFC power of 20 kW and 50 kW, respectively.

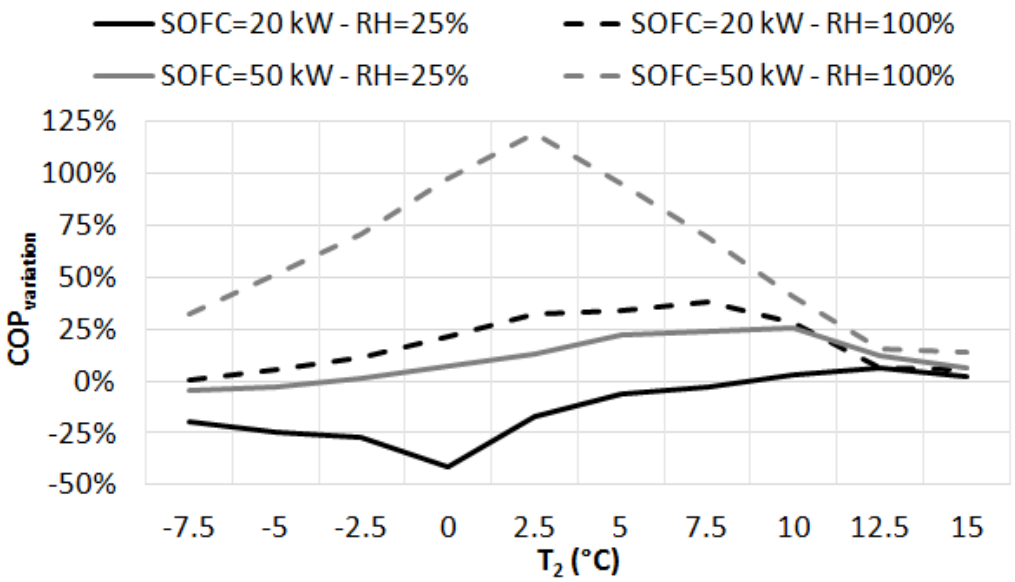


Figure 18. COP variation varying the external inlet air temperature for four very different cases in terms of SOFC nominal power and air relative humidity.

The main conclusion of this reasoning is that the adiabatic mixer has a positive effect on the heat pump COP when outdoor air has a high relative humidity and when SOFC electric power is high. When SOFC power is 50 kW, COP of heat pump is always improved by the mixer. If a 20 kW SOFC is used, then the present system has a higher COP only when relative humidity of the inlet air is close to 100 %.

Figure 18 shows another interesting aspect of the system in which the higher air relative humidity and SOFC power, the lower air temperature at which maximum $COP_{variation}$ occurs. If inlet air temperature is above 12.5 °C, the adiabatic mixer is not useful at all.

1.4.8 RESULTS - PRIMARY ENERGY SAVING

As previously mentioned, *%PES* is defined to quantify energy saving from the proposed system and compared to the traditional solution (system). Figure 19 depicts *%PES* as a function of the outdoor air temperature for the same four different cases as in Figure 18 (very different cases but representative). The proposed system allows a primary energy saving in the range of 37.5 % – 45 %. The system with relatively small SOFC power presents lower *%PES* compared to the case with relatively high SOFC power, only for temperature below 2.5 °C and when relative humidity is low (25 %). Such critical value of air temperature decreases to 0.5 °C when humidity is high (close to 100 %).

Note that the higher the relative humidity, the higher *%PES* in the given temperature range of T_2 (dotted lines are always above continuous lines in Figure 19). It is also worth to note that primary energy saving depends also on the partial load operation of SOFC. This is due to variation of power ratio, and thereby the efficiency of SOFC.

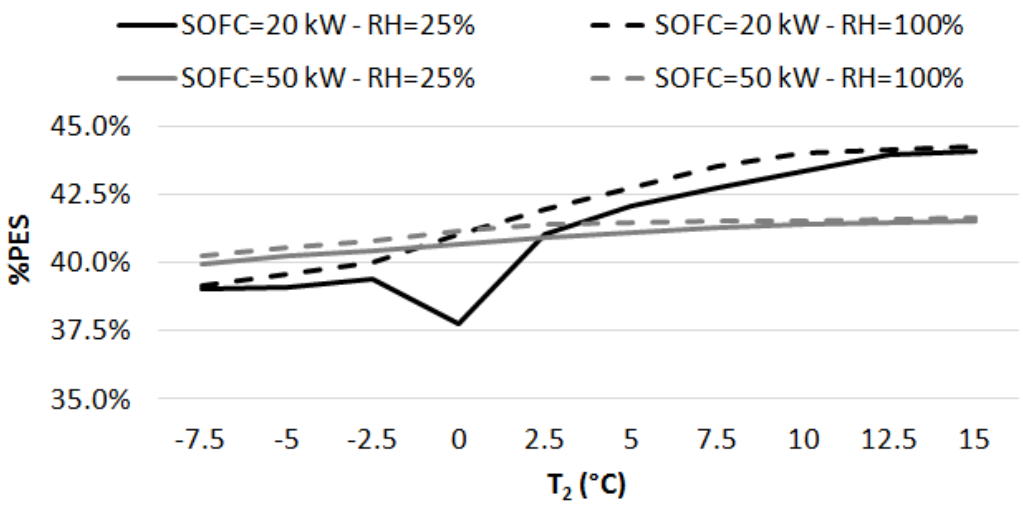


Figure 19. Primary energy saving varying the external inlet air temperature for four very different cases in terms of SOFC nominal power and air relative humidity.

1.4.9 CONCLUSIONS

A cogeneration system based on SOFC fed by natural gas (producing both electricity and heat) and ASHP with an advanced heat recovery system (to increase overall heat production) is analysed and proposed to enhance overall efficiency of the system on energy generation. Outdoor air entering the evaporator of heat pump is mixed with exhausted gases from the SOFC plant with the aim of increasing evaporator temperature and thereby reducing possibility of freezing. Such conditions allow increasing the coefficient of performance for the heat pump.

Simulations are performed varying the external air temperature, air humidity and SOFC nominal power. Thermodynamic analysis shows that in some cases the effect of mixing the exhausted gases with air is negative: when SOFC electric power is lower in comparison to its nominal power (50 kW) and/or inlet air has a low relative humidity, COP decreases up to 35 %. Instead, COP increases up to 100 % when SOFC electric power is close to its nominal, and/or inlet air has a high relative humidity. A comparison based on primary energy consumption between the system proposed here and a traditional one with separate production (electricity demand is covered by the national grid and heat demand is produced separately by a boiler) proves that significant savings can be achieved (between 37.5 % and 45 %).

Results show that ASHP performance could be increased considerably during cold season for climates with high relative humidity. Such results quantify the coefficient of performance and primary energy saving of the SOFC-ASHP integrated system.

1.5 RSOC - SOEC

1.5.1 INTRODUCTION

Hydrogen will play a key role into the next energy systems: analyses and studies have been already performed ([34]–[38]) to verify the state-of-art, the production costs, and the uses as energy vector for the transport sector and as energy storage. As a matter of fact, hydrogen is proposed for:

- *Decarbonized transport sector*: hydrogen as the energy vector for transport when fuel cells instead of combustion engines are proposed, with the aim of reducing pollution and CO_2 emission;

- *Management of high energy price fluctuation*: hydrogen can be used to reduce the cost fluctuation of energy: it can be produced when low costs occur, and then converted into electricity when prices are high;

- *Management of energy system with high share of renewable energy*: mismatching between production and demand of energy can be frequent in energy systems with high share of renewable energy. Energy sources such as solar or wind are highly related to weather conditions and are not programmable. As a matter of fact, in some periods the energy production is higher than the energy demand, in others vice versa. Hydrogen is proposed as an energy storage, as it is produced when a surplus of energy production is available and it is consumed when a deficit of energy is present ([35], [36]).

Hydrogen is actually used in chemical industry, but according to [38] it is mainly produced by the reforming of natural and refinery gases (48%), as a by-product of chemical production (30%), and by coal gasification (18%). Only a small amount is produced with water electrolysis. The production of hydrogen from hydrocarbons and/or coal is considered not suitable in the long run because CO_2 is produced. Water electrolysis using renewable energy is proposed as main alternative to decrease greenhouses emission. Water electrolysis is an electrochemical reaction where water is divided into hydrogen and oxygen using thermal and electrical energy:

$$2H_2O \rightarrow 2H_2 + O_2 \qquad \text{Eq. 21}$$

The overall energy request of the reaction (ΔH) is partially supplied by heat (ΔQ) and partially in an electrical manner by changing the Gibbs energy (ΔG) (Eq. 22):

$$\Delta H = \Delta Q + \Delta G \qquad \text{Eq. 22}$$

The theoretical minimum cell voltage (V_{rev}) of the electrolyser could be calculated under the assumption that an external heat source supplies the thermal energy requested. It is directly related to the Gibbs free energy (ΔG) according to the equation Eq. 23:

$$V_{rev} = \frac{\Delta G}{nF} \qquad \text{Eq. 23}$$

Where n is the number of electrons transfers per reaction (n=2) and F represents the Faraday's constant (96485 C/mol). Thermoneutral voltage is defined, instead, under the assumption that electrolysis takes place without heat integration, consequently the overall energy demand (including heat) is provided electrically. Thermoneutral voltage (V_{tn}) is defined as (Eq. 24):

$$V_{tn} = \frac{\Delta H}{nF} \qquad \text{Eq. 24}$$

Figure 20 represents how electrical, thermal and total energy demands vary with temperature. Firstly, V_{rev} decreases slightly with the increase of the temperature of the reaction in a range of 1.25 V - 0.91 V between 0 °C and 1000 °C. Thermoneutral is approximatively 1.47-1.48 V if the cell works below 100 °C, and 1.26-1.29 V if the temperature range is between 100 – 1000 °C.

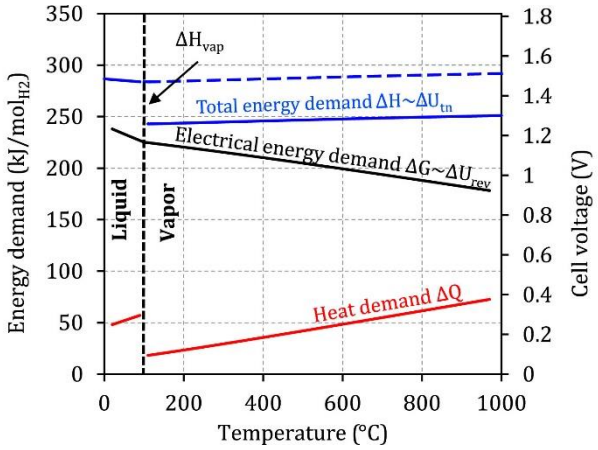


Figure 20. Electrical (ΔG), thermal (ΔQ) and total (ΔH) energy demand variation on temperature ([39])

The main water electrolysis technologies currently available are proton exchange membrane electrolysis (PEMEL or PEM), alkaline electrolysis (AEL) and solid oxide electrolysis (SOEL). Differences occur on the electrolyte of the cell and, consequently, on working temperature, specific energy demand and efficiency on energy conversion.

The alkaline electrolyser has a 25-30% aqueous KOH solution as electrolyte, and the electrolytes are immersed and separated by a diaphragm. It is considered a low temperature electrolyser working at 60 – 90 °C. Partial reactions at the electrodes are represented in Figure 21 .

Anode	Cathode
$2OH^- \rightarrow H_2O + \frac{1}{2}O_2 + 2e^-$	$2H_2O + 2e^- \rightarrow 2H_2 + 2OH^-$

Figure 21. Partial reactions in alkaline electrolyser

Figure 22 represents how the system works: water is fed at the anode side where hydrogen is produced, two drums store the produced gases (one for hydrogen and one for oxygen) with the electrolyte working as gas-liquid separator. External cooling system is required to dissipate heat produced during electrolysis. Gas quality after drying is typically in the range of 99.5–99.9% for H_2 and 99–99.8% for O_2; catalytic gas purification could increase gas purity to 99.999%. Commercial electrolysis stacks work at 1.25 – 2.1 V (varying on current density) and a current density of 0.4 A/cm^2, reaching a rated efficiency of 63–71% on hydrogen LHV and a specific energy demand of 4.2–4.8 kWh/Nm^3 [39] (Table 7).

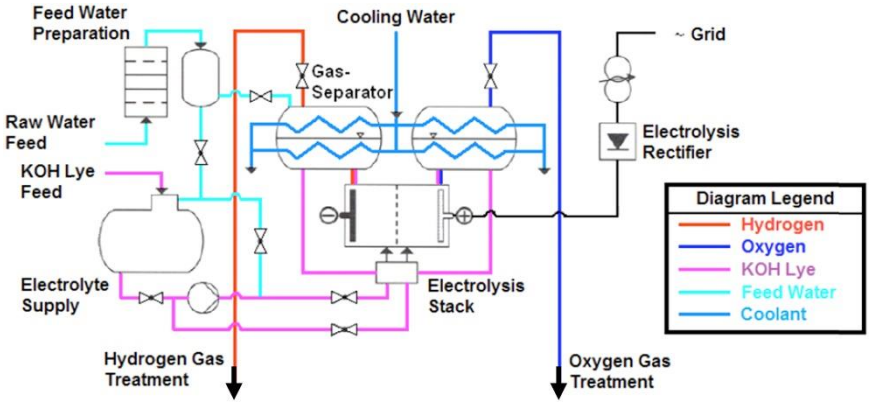


Figure 22. Alkaline electrolyse system [39]

Proton Exchange Membrane Electrolyser (PEMEL) was introduced in 1960s by General Electric. Electrodes are directly mounted on a proton exchange membrane (usually Nafion © membrane). PEMEL is considered also a low temperature electrolyser, as it works at 60-90 °C.

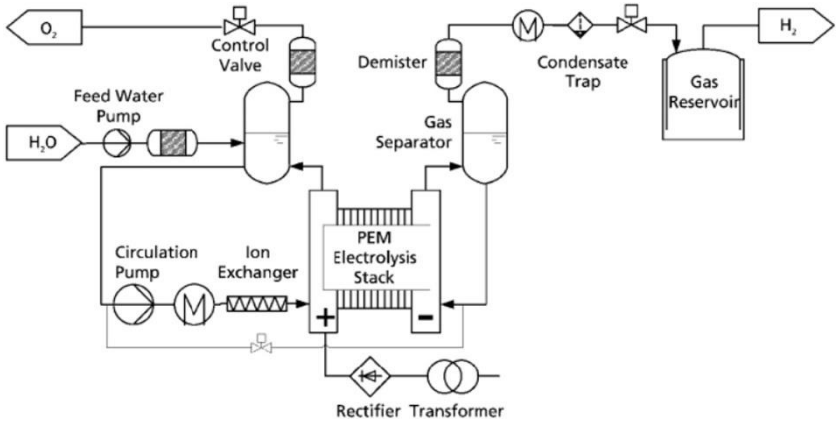


Figure 23. PEMEL system ([39])

Figure 23 represents a PEMEL system: water is supplied at the anode side where oxygen is produced, meanwhile hydrogen is available at cathode side. The following partial reactions take place into the electrolyser (Figure 24):

Anode	Cathode
$H_2O \rightarrow \frac{1}{2}O_2 + 2H^+ + 2e^-$	$2H^+ + 2e^- \rightarrow H_2$

Figure 24. Partial reaction of PEMEL

The corrosive acidic regime provided by the proton exchange membrane requires the use of noble metal catalysts like iridium for the anode and platinum for the cathode. Compared to alkaline electrolyser, PEMEL has a very low cross-permeation: H_2 with a 99.99% purity is available after drying. One of the featured characteristics of PEMEL compared to AEL is the compact module design and the higher current density operation. Consequently, a high-pressure operation on the cathode side (H_2) could be achieved: hydrogen can be produced at higher pressure than AEL. This is a key feature if compressed hydrogen is required, energy requested to compress hydrogen after electrolysis is higher than the one requested by compression of water before electrolysis. The cell voltage in PEMEL stack is 1.6 – 2.5 V with a current density of 1 - 2 A/cm^2; stack efficiency is reported to be 60–68% on hydrogen LHV with an energy demand of 4.4–5.0 kWh/Nm^3 [39] (Table 7).

The solid oxide electrolyser works at higher temperature compared to AEL and PEMEL, around 700 – 900 °C. High working temperature requires that electrolyte is based on solid oxide. Figure 25 represents a simplified layout of a system: electrolyser is based on planar cells, and water is provided as steam at high temperature. Higher working temperature implies a lower electricity consumption as showed in **Errore. L'origine riferimento non è stata trovata.**: part of the energy is given by heat.

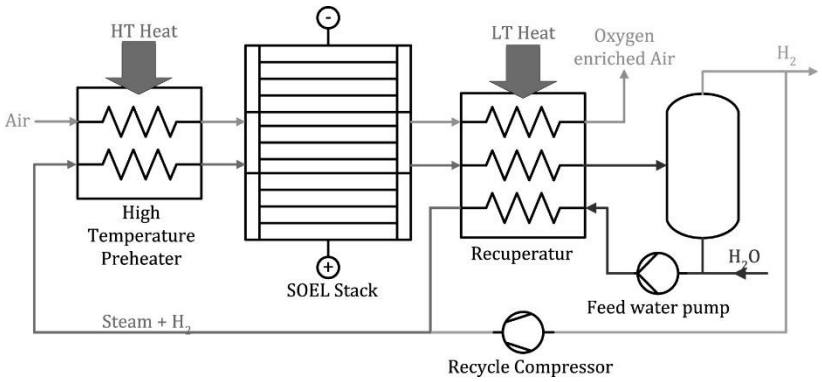


Figure 25. Solid oxide electrolyser system ([39])

Partial reactions at the electrodes are (Figure 21 and Figure 26):

Anode	Cathode
$O^{2-} \rightarrow \frac{1}{2}O_2 + 2e^-$	$H_2O + 2e^- \rightarrow H_2 + O^{2-}$

Figure 26. Partial reactions in solid oxide electrolyser

AEL and PEMEL are considered mature technologies, and commercial system are currently available at megawatt power scale. Instead, solid oxide electrolysers are still under development with few commercial systems available within kilowatt power scale. It is reported an efficiency of 81 % on hydrogen LHV with an energy demand of 3.7 kWh/Nm^3 [39] (Table 7).

Type of electrolyte	Nominal power	Energy demand	η_{LHV}
Proton Exchange Membrane (PEMEL)	0.01 – 2 MW	4.4 – 5 kWh/Nm^3	60 % - 68%
Alkaline (AEL)	0.03 – 6 MW	4.2 – 4.8 kWh/Nm^3	63 % - 71 %
Solid Oxide Electrolysis (SOEL)	2.2 kW	3.7 kWh/Nm^3	98 % 84.6 % (including water evaporation)

Table 7. Overview of commercial electrolysis system based on ref. [39], each technology is divided by nominal power, energy demand and efficiency on hydrogen production on LHV

Nowadays there is a great interest in SOEL thanks to its reversibility: reversible solid oxide cells (RSOC) can be used both as electrolyser (solid oxide electrolyser cells – SOEC - if steam and electricity are provided) and as fuel cell (solid oxide fuel cell - SOFC - if hydrogen and oxygen are available). Thanks to this reversibility, RSOC could have a key role in the future energy systems where fluctuations on energy production occur and hydrogen is used as energy storage. In the next section it is proposed a RSOC system as polygeneration system: heat, electricity and hydrogen are produced varying how many cells work as fuel cells and as electrolyser and, consequently, varying H/P ratio on energy generation. If H/P varies, it is possible to decrease the mismatching on energy request and, consequently, increasing efficiency because energy losses are avoided. Hydrogen is a noble by-product of the system and could be used as fuel or, alternatively, for transport. Firstly, the model of the solid oxide cell electrolyser (section 1.5.2) and of the overall system (section 1.5.3) are analysed. Successively, an interaction of SOEC and SOFC is proposed (sections 1.5.4 and 1.5.5) to match H/P of energy demand and production. Finally (section 1.6), a re-vamping of the energy generation system of an industrial facility (a paper mill) using RSOC is proposed and analysed.

1.5.2 SOEC CELL MODEL

In SOEC, water (H_2O) is divided into hydrogen (H_2) and oxygen (O_2) using not only electricity, but also heat in order to decrease the electricity consumption. The SOEC system simulated in this work is based on the model presented in [14], [40] and [41]–[43]. The molar production of H_2 ($\dot{m}_{H_2,out}$) is related to current value, consequently (Eq. 25):

$$\dot{m}_{H_2,out} = \frac{N_{stack} * N_{cell} * A_{cell} * J}{2F} \qquad \text{Eq. 25}$$

Where N_{stack} is the number of stacks, N_{cell} is the number of cells per each stack, A_{cell} is the cell area (m^2), J is the current density (A/m^2) and F is the Faraday constant (96485.34 C/mol). The molar

production of O_2 ($\dot{m}_{O_2,out}$) (Eq. 26) and the residual H_2O ($\dot{m}_{H_2O,out}$) (Eq. 27) can be calculated from the water inlet ($\dot{m}_{H_2O,in}$):

$$\dot{m}_{O_2,out} = \dot{m}_{H_2O,in} - \frac{1}{2}\dot{m}_{H_2,out} = \dot{m}_{H_2O,in} - \frac{N_{stack} * N_{cell} * A_{cell} * J}{4F} \quad \text{Eq. 26}$$

$$\dot{m}_{H_2O,out} = \dot{m}_{H_2O,in} - \dot{m}_{H_2,out} = \dot{m}_{H_2O,in} - \frac{N_{stack} * N_{cell} * A_{cell} * J}{2F} \quad \text{Eq. 27}$$

The SOEC power (P_{SOEC}) is defined as (Eq. 28):

$$P_{SOEC} = N_{stack} * N_{cell} * E_{cell} * A_{cell} * J \quad \text{Eq. 28}$$

where E_{cell} is the cell voltage (V). The minimum electrical work is determined by Nernst potentials but also irreversibilities occur such as activation (ΔE_{act}), polarization (ΔE_{conc}) and ohmic resistance (ΔE_{ohm}) (Eq. 29):

$$E_{cell} = E_{Nernst} + \Delta E_{act} + \Delta E_{ohm} + \Delta E_{conc} \quad \text{Eq. 29}$$

The Nernst potential and the polarization (activation, ohmic and concentration) are calculated as explained in [13], [42] meanwhile the diffusion coefficient is approximated using the kinetic theory and Chapman-Enskog theory [43]. The pressure at the outlet is defined as (Eq. 30):

$$p_{ca\_out} = p_{ca,in} * (1 - dp_{ca}) * p_{an_{out}} = p_{an,in} * (1 - dp_{an}) \quad \text{Eq. 30}$$

where dp_{ca} and dp_{an} are the pressure drops at the cathode and at the anode side respectively; p_{ca_out} , $p_{an_{out}}$ and $p_{an,in}$ are the pressure at the cathode outlet, the anode outlet, and the anode inlet respectively.

1.5.3 SOEC System Proposed

SOECs require both electricity and heat to produce hydrogen. In particular, the latter is necessary to generate high-temperature steam as a reactant and, on the other hand, as a cooling vector for the compressed hydrogen in order to reduce the power for compression. Consequently, SOEC system is mainly composed by SOEC stacks and counter-flow heat exchangers.

Figure 27 represents the proposed system: HEX1 and HEX2 are two heat exchangers that cool hydrogen and oxygen, respectively, after they are produced within the SOEC, while preheating water to obtain steam at the required working condition. HEX3 uses exhaust gases from SOFC to heat the steam up to the SOEC working temperature. The produced hydrogen and oxygen are discharged after SOEC stack at 750 °C, while inlet water has a temperature of 25 °C: heat exchanger are used to cool products (hydrogen and oxygen) to pre-heat water. Without heat recovery performed by HEX1 and

HEX2, HEX3 would exchange a higher thermal power, and it would require a higher exhaust gases flow rate to reach operating temperature, thus increasing the heat demand by SOEC. The SOEC system proposed is simulated with the tool DNA from DTU ([14], [40]). Table 8 resumes the main parameters of the system.

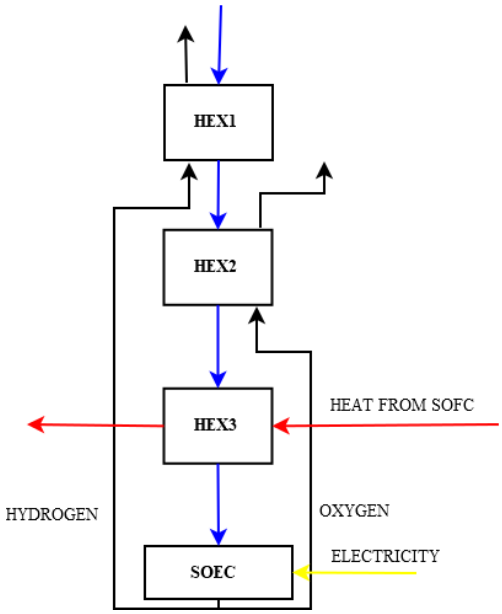


Figure 27. SOEC system layout. Water (blue line) is heated by using produced gases (black lines, hydrogen and oxygen) and exhaust gases from SOFC (red line), in order to obtain steam at operating temperature by means of heat exchangers (HEX1, HEX2 and HEX3). Electricity (yellow line) is needed to perform electrolysis in the SOEC stack.

Table 8. SOEC system parameters.

Parameter	Value
Stack nominal electric power	100 kW
Hydrogen flow (at 90 % of nominal electric power)	2.63 kg/h
Heat request from SOEC (at 90 % of nominal electric power)	15 kW
Working temperature	750 °C
Working pressure	7.01 bar
Pressure drop	0.05 bar

1.5.4 SOFC/SOEC INTERACTION

Different industrial sectors have different heat to power ratios: Figure 2 summarizes how the ratio varies compared with different cogeneration technologies. Reference [3] proposes these technologies: internal combustion engine (ICE), steam turbine, gas turbine, micro turbine and fuel

cell. Each technology has different H/P ratio and electrical capacity: steam turbine has the highest H/P ratio (10-14.3), while fuel cell features the lowest (0.5-1). Electrical capacity is the highest for steam and gas turbines as well (until several hundreds of MW). None of the previous technologies achieves a H/P ratio between 2 and 10 even if many industrial sectors have typical H/P consumption in that range. If a factory has a H/P on energy demand close to 6, none of the cogeneration technologies previously mentioned match the request. Different approaches could be applied, depending on the operation strategy, energy cost and other parameters:

1 Cogeneration system with H/P below 6 (such as ICE or gas turbine): mismatching between H/P of energy demand and energy production could be adjusted with a heat integration system. This would increase heat production or electricity export to the grid. If this solution is adopted, heat integration system (mainly using boiler) is usually chosen because heat demand is usually variable and revenues on electricity export to national grid could not cover electricity generation costs;
2 Cogeneration system with H/P over 10 (such as steam turbine): mismatching between H/P could be adjusted by increasing heat demand (such as district heating as presented in [44]) or dissipating unused heat or importing electricity from grid.

Integration heating system and (in particular) unnecessary heat dissipation could decrease efficiency of energy generation. The main priority for the industry is not only energy efficiency but also cost reduction, that is related to purchase, installation, operative and maintenance costs of the energy generation technology chosen.

RSOC is proposed as innovative cogeneration system for industry because cells can be used to produce electricity and heat (operation as SOFC) or to produce hydrogen consuming electricity and heat (operation as SOEC). Varying the proportion between SOFC and SOEC it is possible to modify H/P ratio of the energy generation. Advantages of a RSOC system are:

- High versatility on fuels: as previously mentioned ethanol [8]–[10], methanol [8]–[10], ammonia [8]–[10], DME [8]–[10], woodchips gasification [11], biomass [12] or municipal waste gasification [13] could be used, it requires only to change fuel pre-reformer. According to [3] only steam turbines have a higher versatility on fuel;
- SOFC and RSOC systems have high modularity, so the scale of the plant can vary from kilowatt to megawatt;
- Heat is produced using heat recovery from SOFC exhausted gases that have a temperature higher than 600 °C, while industrial processes use heat at temperature generally lower than 150 °C (Figure 10);
- Hydrogen is also a product of the RSOC system, it could be re-used as fuel for SOFC or for other uses (exported to natural gas grid, storage and distribution for transports etc).

Figure 28 represents the proposed SOFC-SOEC system (RSOC): SOFC consumes fuel to produce electricity and heat, SOEC consumes a part of electricity and heat available from SOFC to obtain hydrogen and oxygen from water. Unused heat from SOFC can be used to produce steam and/or heat

directly for processes. The aim is to vary the heat-to-power ratio of the system (i.e., the ratio between heat and electricity generated by the system, H/P_{RSOC}) by varying the proportion between the stacks used as SOFC and the ones operating as SOEC, in order to match the energy profiles of the user (a paper mill) at supply and demand sides. If only SOFCs are used, the H/P ratio of the system is equal to the H/P of the SOFC. Instead, if a combination of SOFCs and SOECs are used, a part of the electricity and heat generated by SOFCs is consumed by SOECs to electrolyse water and convert it into hydrogen and oxygen.

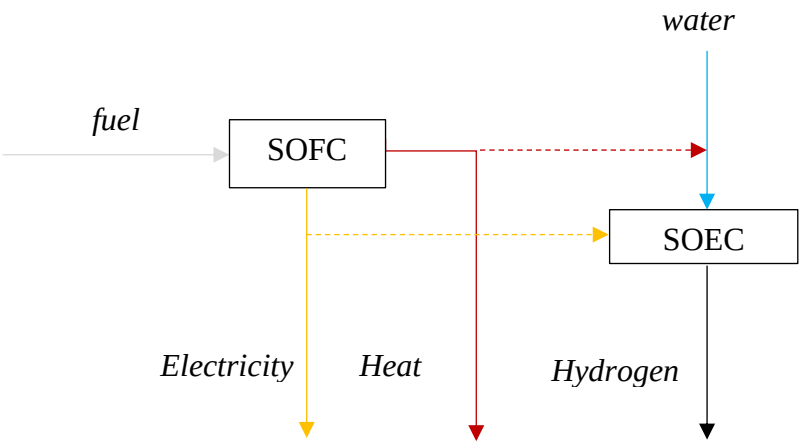


Figure 28. SOFC-SOEC system proposal

1.5.5 MATHEMATICAL MODEL

A mathematical model of the interaction between SOFC and SOEC is proposed in order to define how H/P ratio of energy generation varies with different configurations. n_{RSOC} is defined (Eq. 31) as the ratio between SOEC electricity consumption (P_{SOEC}) and SOFC electricity generation (P_{SOFC}):

$$n_{RSOC} = \frac{P_{SOEC}}{P_{SOFC}} \qquad \text{Eq. 31}$$

H_{RSOC} and P_{RSOC} are the heat and electricity generation of the overall system, respectively. These variables are defined as the difference between SOFC generation (subscript *SOFC*) and SOEC consumption (subscript *SOEC*) in terms of heat (H) (Eq. 32) and electricity (P) (Eq. 33):

$$H_{RSOC} = H_{SOFC} - H_{SOEC} \qquad \text{Eq. 32}$$

$$P_{RSOC} = P_{SOFC} - P_{SOEC} \qquad \text{Eq. 33}$$

Then n_{RSOC} is related to $(H/P)_{RSOC}$, $(H/P)_{SOFC}$ and $(H/P)_{SOEC}$. Firstly, heat-to-power ratio of RSOC system is defined (Eq. 34) considering energy production of SOFC and consumption of SOEC (Eq.

32 and Eq. 33 respectively). An expression that only uses H/P ratios as variables is obtained to relate electricity consumption of SOEC and electricity production of SOFC (Eq. 37).

$$(\frac{H}{P})_{RSOC} = \frac{H_{SOFC} - H_{SOEC}}{P_{SOFC} - P_{SOEC}} \quad \text{Eq. 34}$$

$$(\frac{H}{P})_{RSOC} = \frac{(\frac{H}{P})_{SOFC} * P_{SOFC} - (\frac{H}{P})_{SOEC} * P_{SOEC}}{P_{SOFC} - P_{SOEC}} \quad \text{Eq. 35}$$

$$(\frac{H}{P})_{RSOC} = \frac{(\frac{H}{P})_{SOFC} - (\frac{H}{P})_{SOEC} * \frac{P_{SOEC}}{P_{SOFC}}}{1 - \frac{P_{SOEC}}{P_{SOFC}}} \quad \text{Eq. 36}$$

$$(\frac{H}{P})_{RSOC} = \frac{(\frac{H}{P})_{SOFC} - (\frac{H}{P})_{SOEC} * n_{RSOC}}{1 - n_{RSOC}} \quad \text{Eq. 37}$$

n_{RSOC} is defined (Eq. 38) directly from $(H/P)_{RSOC}$, $(H/P)_{SOFC}$ and $(H/P)_{SOEC}$ as:

$$n_{RSOC} = \frac{(\frac{H}{P})_{RSOC} - (\frac{H}{P})_{SOFC}}{(\frac{H}{P})_{RSOC} - (\frac{H}{P})_{SOEC}} \quad \text{Eq. 38}$$

Figure 29 shows how the heat-to-power ratio of RSOC system ($(H/P)_{RSOC}$) varies with n_{RSOC}. SOEC consumes electricity and heat to electrolyse water, and energy is provided by SOFC: the higher the SOEC utilisation, the lower the electricity and heat available from SOFC. $(H/P)_{RSOC}$ represents the ratio between electricity and heat available from the RSOC system: the higher the fraction of stacks operating as SOEC, the higher the $(H/P)_{RSOC}$. $(H/P)_{RSOC}$ increases when the proportion shifts towards SOEC presence, because SOEC consumes proportionally more electricity than heat from SOFC. n_{RSOC} is useful to analyse how to match $(H/P)_{SOFC}$ and heat-to-power ratio required by the user $(H/P)_{USER}$: if $(H/P)_{USER}$ is higher than $(H/P)_{SOFC}$, an increase of SOEC utilisation is needed to increase $(H/P)_{RSOC}$.

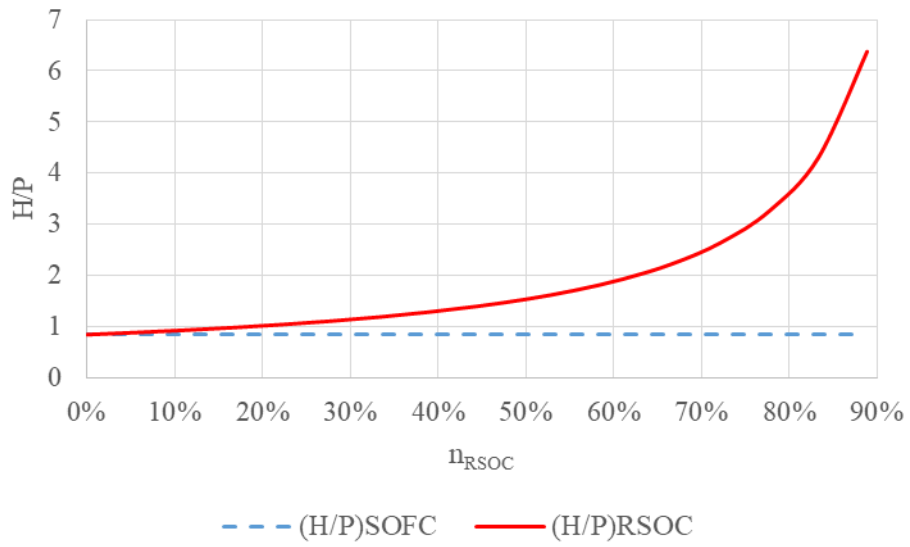


Figure 29. RSOC system heat to power ratio versus n_{RSOC}

Given P_{RSOC}, which is known because it is the electricity required by the RSOC system, P_{SOFC} and P_{SOEC} can be defined through n_{RSOC}. Considering the previous equations, P_{SOFC} and P_{SOEC} can be obtained respectively by Eq. 39 and Eq. 40:

$$P_{SOFC} = \frac{1}{1 - n_{RSOC}} * P_{RSOC} \qquad \text{Eq. 39}$$

$$P_{SOEC} = \frac{n_{RSOC}}{1 - n_{RSOC}} * P_{RSOC} \qquad \text{Eq. 40}$$

The RSOC number of stacks is directly related to its nominal power and, consequently, to the nominal power of both SOFC and SOEC. The parameter $P_{MAX,RSOC}$ is defined as the sum of SOFC electric generation and SOEC electric consumption, so the maximum electric power of the RSOC system (Eq. 41) and, accordingly, the number of stack required can be calculated.

$$P_{MAX,RSOC} = P_{SOFC} + P_{SOEC} \qquad \text{Eq. 41}$$

Eq. 42 combines Eq. 39, Eq. 40 and Eq. 41 to obtain $P_{MAX,RSOC}$ (maximum electric power of RSOC), by relating n_{RSOC} (ratio between SOEC and SOFC) and P_{RSOC} (electricity production of RSOC system).

$$P_{MAX,RSOC} = \frac{1 + n_{RSOC}}{1 - n_{RSOC}} * P_{RSOC} \qquad \text{Eq. 42}$$

Eq. 37 and Eq. 42 are useful to size the RSOC system, which is the novel element proposed in this study. In particular:

1. $(H/P)_{RSOC}$ and P_{RSOC} are known due to energy request of the user, while $(H/P)_{SOFC}$ and $(H/P)_{SOEC}$ are gathered from technical datasheet and/or simulation. Eq. 37 defines the ratio between SOEC and SOFC power (n_{RSOC}) by using these parameters;
2. Eq. 42 defines the size of the RSOC system ($P_{MAX,RSOC}$) and, consequently, the number of stacks as a function of the ratio between SOEC and SOFC (n_{RSOC}), and electricity generation of RSOC system (P_{RSOC}).

For the sake of clarity and a better comprehension of the entire system, each variable used in the previous equations is also represented in Table 9, which shows the structure of the novel system.

COMPONENT	VARIABLE	DESCRIPTION
RSOC	H_{RSOC}	Heat production of RSOC system
	P_{RSOC}	Electricity production of RSOC system
	$(H/P)_{RSOC}$	Heat to power ratio of RSOC system
	n_{RSOC}	Ratio between electricity consumption of SOEC and electricity production of SOFC
	$P_{MAX,RSOC}$	Sum of both electric power of SOFC and electric power of SOEC
SOFC	H_{SOFC}	Heat production of SOFC system
	P_{SOFC}	Electricity production of SOFC system
	$(H/P)_{SOFC}$	Heat to power ratio of SOFC system
SOEC	H_{SOEC}	Heat consumption of SOEC system
	P_{SOEC}	Electricity consumption of SOEC system
	$(H/P)_{SOEC}$	Heat to power ratio of SOEC system consumption

Table 9. List of RSOC, SOFC and SOEC variables in order to define the mutual interaction.

Figure 30 compares H/P of the proposed system varying n_{RSOC} and other cogeneration technologies: gas turbine, ICE, micro gas turbine and steam turbine. It is possible to appreciate that RSOC can cover H/P range between 2 and 10 in which no cogeneration system is available, even if the proposed system in that range mainly consumes fuel to produce hydrogen: n_{RSOC} is higher than 50%, consequently it works more in SOEC than SOFC mode.

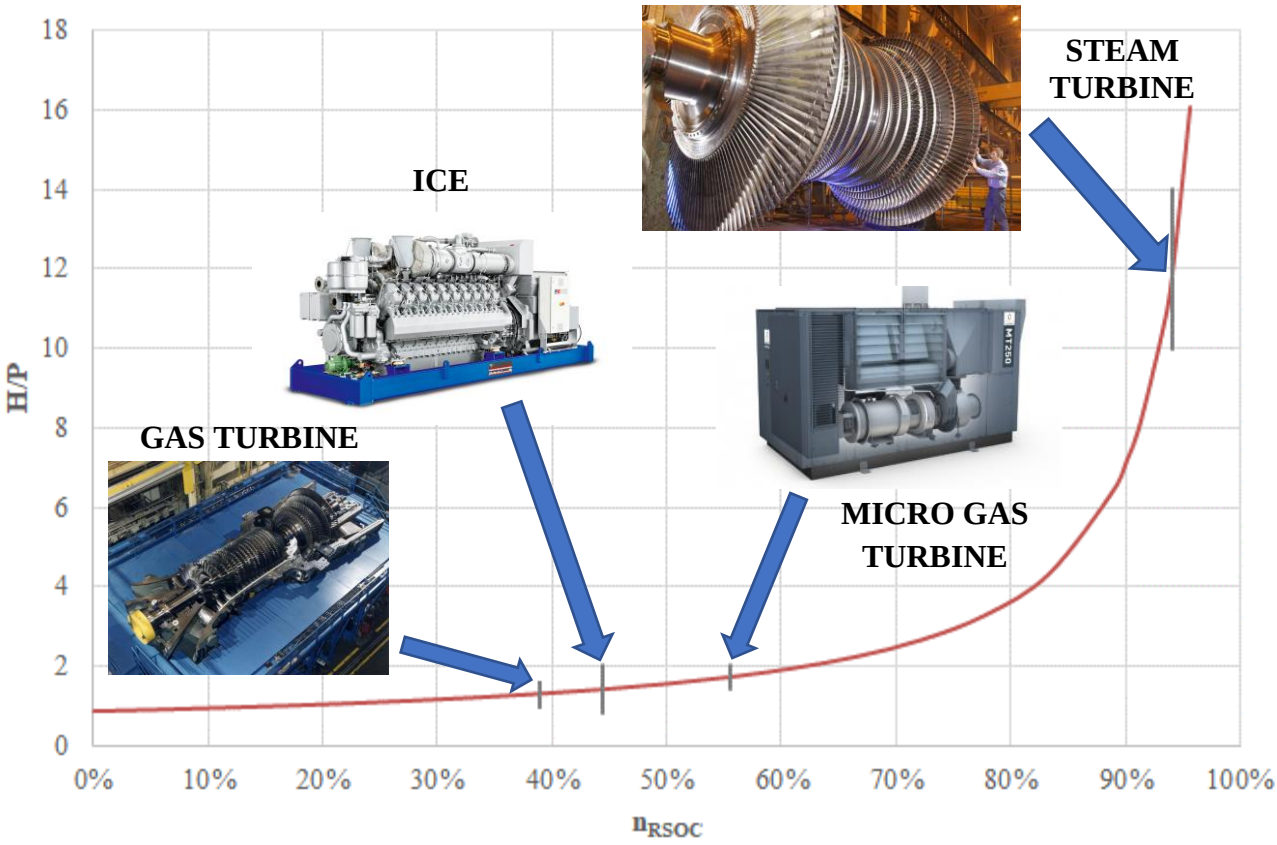


Figure 30. H/P_{RSOC} variation on n_{RSOC} compared with H/P of other cogeneration technologies (gas turbine, ICE, micro gas turbine, steam generator) (images from [45]–[48]*)*

1.6 ENHANCEMENT OF ENERGY GENERATION SYSTEM IN INDUSTRIAL FACILITY WITH SOFC/RSOC

1.6.1 INTRODUCTION

Many countries have already implemented subsidies on energy efficiency in generation and utilisation, with the aim of decreasing overall consumption and energy intensity of gross domestic product. Many processes in the industrial sector are highly energy intensive, so that energy efficiency measures and innovative conversion solutions are deeply investigated for both environmental protection [49] and operational costs reduction [50]. Advantages can be evaluated in either purely technical (energy saving) or techno-economic (cost reduction) terms. Typical examples are industrial applications where high energy flows are treated, like metals and plastic manufacturing [51], or wood processing [52]. A number of studies has investigated glass production and the potential for energy performance improvement with different integrated processes [53], [54]. Energy analyses on cast iron foundries have been developed, proposing system improvements to decrease and control energy demand [55]–[57].

The contemporary request of heat and electricity characterizes the pulp and paper industry, which thus opens to innovative process integration for energy efficiency improvement [58], [59]. Focusing on the final step, i.e. paper production in paper mills, two energy vectors are involved: heat to satisfy the need of drying the mid-products, and electrical power to drive the motors. The use of CHP units in paper mills is a known strategy, typically implemented with conventional internal combustion engines or steam turbines [60], [61]. Decarbonisation strategies have been investigated considering the use of renewable biomass or waste from the facility itself as sources, as well as the introduction of carbon capture technologies [62], [63].

In this study, the re-vamping of the energy generation system of a paper mill by means of Reversible Solid Oxide Cells (RSOCs) is proposed. The aim is not only to increase efficiency on energy generation, but also to create a polygeneration system where hydrogen is produced. The study focuses on a real industrial facility, based in Italy, with a production capacity of 60000 t/y of paper. As the main novelty of this study, the adoption of a solid oxide fuel cell / solid oxide electrolyser cell (SOFC/SOEC) system is proposed, substituting of part of the existing set of cogenerators. This introduces an additional hydrogen production as a result of the need to balance the mismatch of heat-to-power ratios between energy production and demand. As a matter of fact, this enlarges the plant activity beyond paper production, exploiting the inherent capability of managing high-temperature flows and bringing it into the field of multi-energy integrated systems, which are expected to become more and more relevant in the clean energy future.

In the field of multi-energy systems, hydrogen has been studied lately as an alternative for both power generation and storage. Canan *et al.*[34] have studied hydrogen production from renewable and non-renewable sources to assess environmental impact, production costs, energy and exergy efficiency of the different methods. Astiaso Garcia *et al.*[35] have surveyed and analysed potential of hydrogen as energy storage systems in EU countries to reduce energy fluctuations and possible negative effects due to an increase of renewable sources share in power generation. Guandalini *et al.*[36] have investigated the use of Power-to-Gas systems coupled with wind farms to improve dispatchability. Castellani *et al.* [37] have studied the use of hydrogen produced with renewable energy for flue gas treatment to produce methane and ammonia, in order to reduce carbon footprint of the process. The main novelty of the multi-energy integrated system proposed in this thesis is that can be used to revamp or to substitute cogeneration systems in industrial facility. Thanks to its flexibility on H/P ratio on energy generation, it decreases the mismatching between energy production and demand. Other cogeneration technologies such as ICE, gas turbine or steam turbine have a lower flexibility because H/P ratio can vary only into a limited range (Figure 30). On the other hand, hydrogen is a sub product: if it is adopted on large scale, the proposed system would introduce a de-centralised hydrogen production scenario, that could be a first stage towards its introduction as energy vector.

1.6.2 OVERVIEW OF THE INDUSTRIAL FACILITY AND ACTUAL ENERGY GENERATION SYSTEM

The industrial facility under consideration is a paper mill located in the North-East of Italy. It produces paper with a weight between 40 g/m^2 and 500 g/m^2, and it is equipped with two paper production lines with an annual capacity of 60000 t/year. Figure 31 represents the scheme of one production line within the paper mill, highlighting the different processes involved. First, pulp is mixed with water and raw materials (e.g., kaolinite, a type of clay) in the headbox. Then, the obtained pre-product is sent to the wet-press section where rollers press it to start removing water from the pulp. Next, in the drying section, heated rollers and felts further decrease water content and thickness of the pre-product down to production specifications. Each production line consumes electricity for motors (each roller has an independent one) and vacuum pumps (felts uses vacuum instead of heat for drying), while heat is required to heat up rollers (steam is generally used). In some cases, natural gas-fired dryers are used within the drying section.

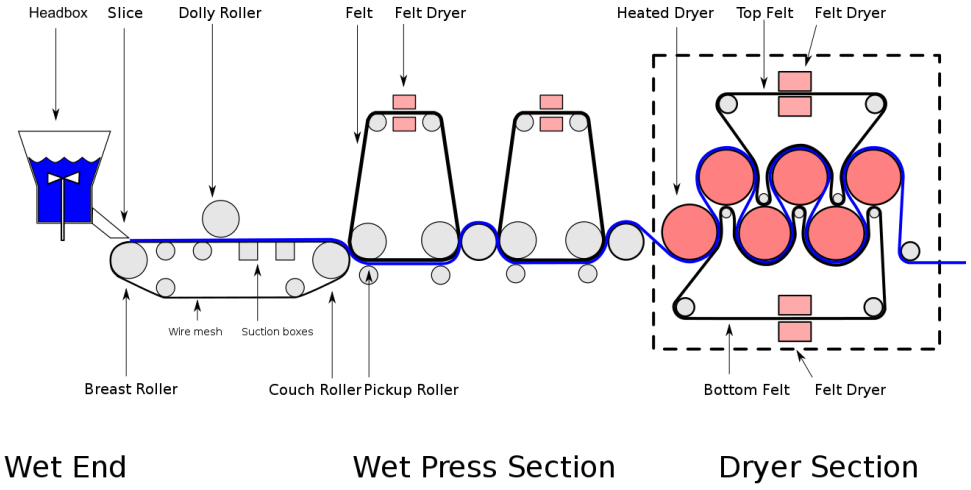


Figure 31. Schematic representation of a paper production line.

After the economic crisis of 2008, the drop in paper request reduced the production rate significantly. Due to the high start-up time (typically half to one day), a paper production line must run continuously, with the workers active on a three-shift basis. In order to match the lower paper request to the plant capacity, the production cycles have been organised in series of some full-load days followed by days off, e.g. each production line is run continuously for 6 days and then it is switched off for 3 days. Figure 32 represents the current organisation of the production. In Case 1, only line 1 works; in Case 2, only line 2 works; in Case 1+2, both lines 1 and 2 are in operation. One of the consequences of this production plan is that the total operating time is uniformly divided into three situations: for one third of the time all the paper production lines work, for another one third only Line 1 works, and for the last one third only Line 2 is in operation. As electricity and steam consumptions are generally variable with time, in this work a representative average value of electricity and heat consumption is considered for each case, extracted from real data.

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9
Line 1									
Line 2									
Case	Case 1	Case 1+2	Case 1+2	Case 2	Case 2	Case 2	Case 1+2	Case 1	Case 1

Figure 32. Production organisation of the paper mill considered.

The considerable amounts of electricity and steam required by the industrial process suggests the adoption of a cogeneration system. In the current configuration, the industrial facility has two cogeneration units. The oldest one is based on a natural gas-fired steam generator (installed in the '70s to substitute the previous coal-fired unit), coupled to two steam turbines (still the original ones first installed in the '40s). The newest cogeneration unit (installed in 2013) features an internal combustion engine (ICE) with heat recovery. Figure 33 represents the overall energy system: both the ICE and the steam turbines discharge steam at the same thermodynamic conditions (235 °C and 280 kPa, as required by the process), and generate electricity. Steam is used in the paper production process mainly to heat up rollers, and it returns to the ICE and the steam generator as condensed water at a temperature of 80 °C and atmospheric pressure. Both the steam generator and the ICE use natural gas as fuel.

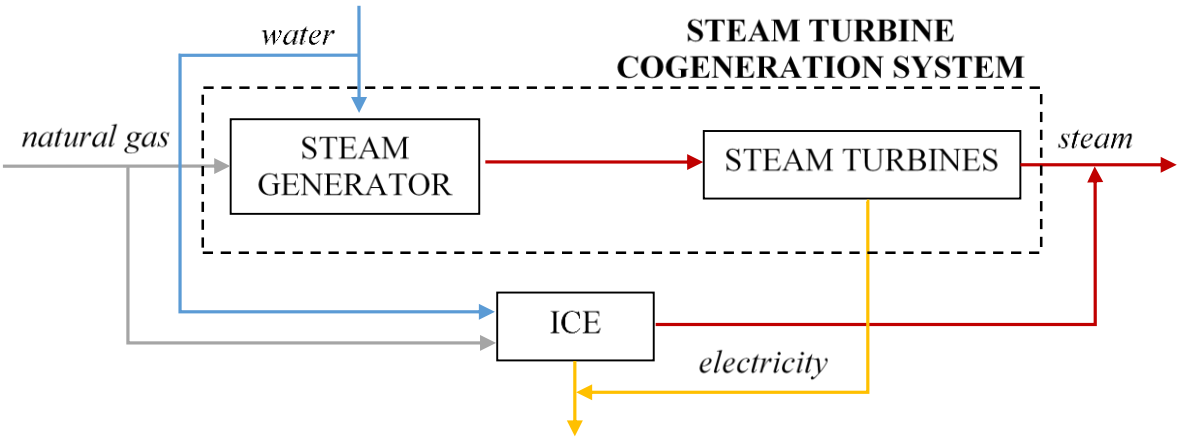


Figure 33. Overview of the cogeneration system with the main equipment and energy flows. Grey flows are natural gas, blue flows are water, red flows are steam, and yellow flows are electricity.

An analysis is performed in order to revamp the current energy system (in particular, the oldest component, that is the steam turbines) with fuel cells that are expected to have higher efficiency. In the next sections both steam turbine and ICE cogenerator are described.

1.6.3 STEAM TURBINE COGENERATION SYSTEM

The steam turbine cogeneration system is composed by a steam generator, two steam turbines, and a thermal user (the paper drying process) that also acts as condenser (Figure 33). The steam generator is fired by natural gas and releases steam at 420 °C and 35 bar (3500 kPa). The steam turbines expand the fluid, thus generating power while decreasing the pressure from 35 bar to 2.8 bar,

which is the pressure required by the paper dryer cylinders. During the expansion process, the temperature decreases from 420 °C to 235 °C. The requested steam mass flow rate varies according to the number of paper machines operating. Steam turbines operation depends on the steam flow rate required for the paper production process: turbine 1 is used when steam generation is between 5 ton/h and 21 ton/h, turbine 2 is used when steam generation is higher. Both steam turbines are more than 50 years old and, even with proper annual maintenance, low performances are expected. In the present work, the steam generator and the turbines are modelled by means of characteristic curves that link fuel demand, electrical generation, and steam mass flow rate. The curves are obtained via mathematical regression from operational data collected in recent years. The information provided by the company about the turbines includes regression curves, so these are used for the analysis directly as given. For the steam generator performance, the curve is obtained from the provided measured values via numerical regression. Figure 34 represents data and curve of natural gas consumption as a function of generated steam mass flow rate (ton/h). The natural gas flow rate (m^3/h) is converted into energy demand (MW) using a reference lower heating value (LHV) of natural gas equal to 9.91 kWh/m^3. Equations in Table 10 describe each component (Eq. 43 for steam generator, Eq. 44 for steam turbine 1 and Eq. 45 steam turbine 2) with a suitable expression, where $\dot{m}_{steam}$ is the mass flow rate of the steam (ton/h), $Fuel_{cons,ST}$ is the natural gas consumption of the steam generator (MW), and P_{ST} is the electric power generated by the turbine (kW).

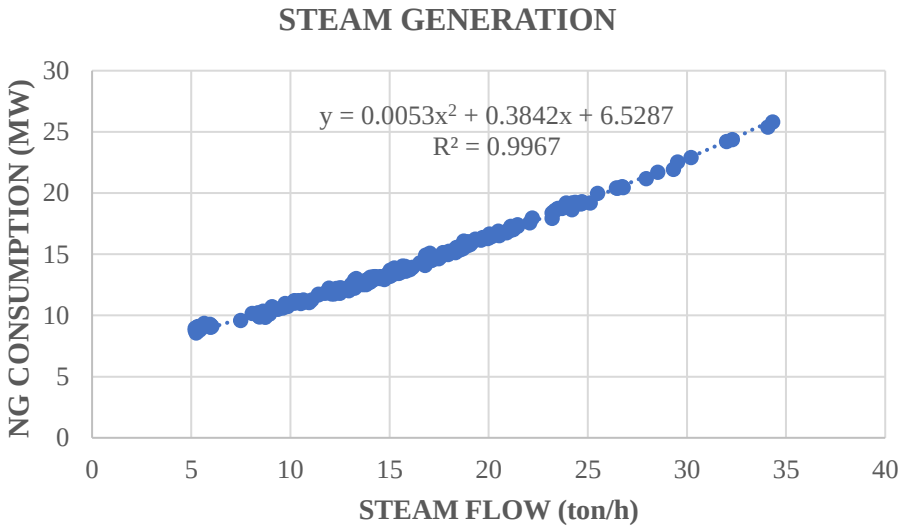


Figure 34. Data sampling of steam generator and regression.

Component	Equation	Condition	
Steam generator	$Fuel_{cons,ST} = 0.0053 * \dot{m}_{steam}^2 + 0.3842 * \dot{m}_{steam} + 6.5287$	5 ton/h < $\dot{m}_{steam}$ < 35 ton/h	Eq. 43
Turbine 1	$P_{ST} = 0.0007 * \dot{m}_{steam} + 6.3908$	5 ton/h < $\dot{m}_{steam}$ < 21 ton/h	Eq. 44

Component	Equation	Condition	
Turbine 2	$P_{ST} = 5.3216e^{0.0007*m_{steam}}$	$\dot{m}_{steam} \geq 21$ ton/h	Eq. 45

Table 10. Mathematic description of each component of the steam turbine cogenerator.

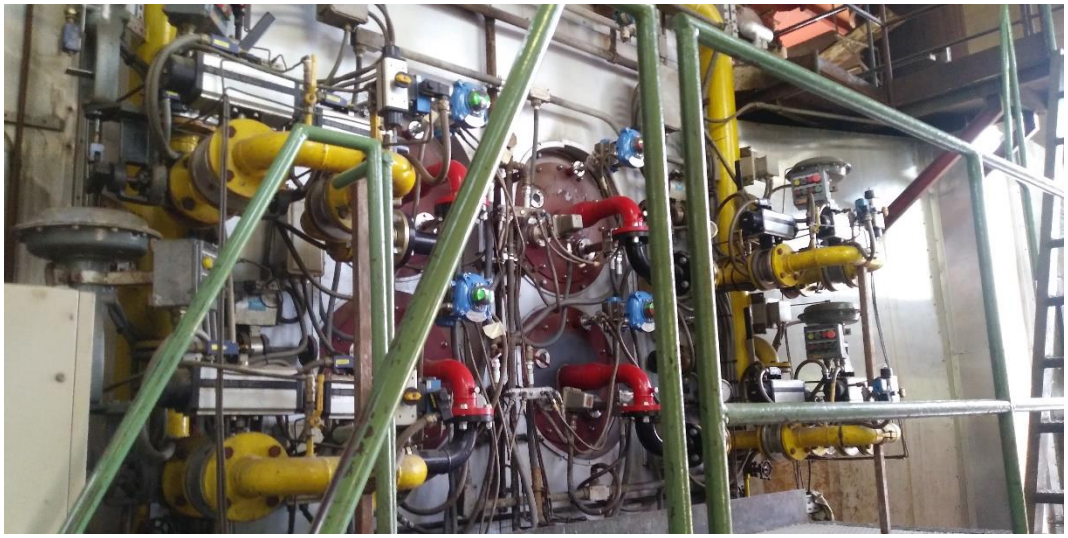


Figure 35. Steam generator, photo of natural gas burners

Figure 36. Photo of the existing steam turbines (turbine 1 in front, turbine 2 behind).

1.6.4 INTERNAL COMBUSTION ENGINE COGENERATOR

In the beginning of 2013, the company decided to increase its energy generation capacity, both electricity and heat, adding an ICE fuelled by natural gas with a nominal electric power of 4.3 MW. Heat recovery from exhaust gases allows producing steam at the same pressure and temperature of

the turbines discharged steam (235 °C, 3.5 bar). It is possible also to recover heat from the engine cooling water and the intercooler, but this heat is available at low temperature, so it cannot be used for steam generation purposes. In some very limited periods of the year (typically during winter), such low-temperature heat is used for space heating and/or for auxiliaries, anyway for most of the time it is dissipated using dry coolers. Table 11 summarises main ICE nominal parameters. The performance is simulated under different operating conditions based on data taken from its technical datasheet. ICE could work at partial load down to 40 % of the nominal electric power.

Parameter	Value
Nominal electric power	4.3 MW
Nominal fuel consumption	10.07 MW
Steam outlet temperature	235 °C
Steam outlet pressure	3.5 bar (350 kPa)
Nominal electric efficiency	42.7%
Nominal thermal efficiency	44.1%

Table 11. ICE nominal parameters.

1.6.5 POLYGENERATION SYSTEM PROPOSED

Reference [2] suggests that paper mills have a H/P ratio of 5. On the other hand, steam turbines have a H/P between 10 and 14, gas turbines between 0.9-1.6, and fuel cells lower than 1 ([3], Table 1). Considering the facility described previously, steam turbines are the units under discussion for modification or substitution in order to improve the energy performance, as they feature very low efficiency. The proposal is to adopts RSOC system both in SOFC and SOEC mode: reversible fuel cells are proposed because SOFCs are used as alternative cogeneration units, meanwhile SOECs recover excess power generation when the heat-to-power ratio impose to over-generate electricity to satisfy the heat demand. This is likely to occur not rarely, due to a SOFC heat-to-power ratio very different from the steam turbine one. The rationale of the proposed concept is the high efficiency of the SOFC systems as well as the already proven capability of the facility to manage high-temperature flows. Moreover, this change allows the paper mill to enlarge its sector of activity, entering into the field of multi-energy systems. Indeed, the expected net hydrogen production could be exploited directly as fuel, thus reducing the natural gas consumption (used as fuel for steam generator, ICE and RSOC working in SOFC mode), or be addressed to an external market, e.g. hydrogen for mobility. Figure 37 represents the novel system proposed with the scheme of the energy flows. Steam turbines, ICE and SOFC produce both electricity and heat for steam production, SOEC consumes a part of SOFC electricity and heat to both produce hydrogen and to match H/P ratio between energy generation and consumption.

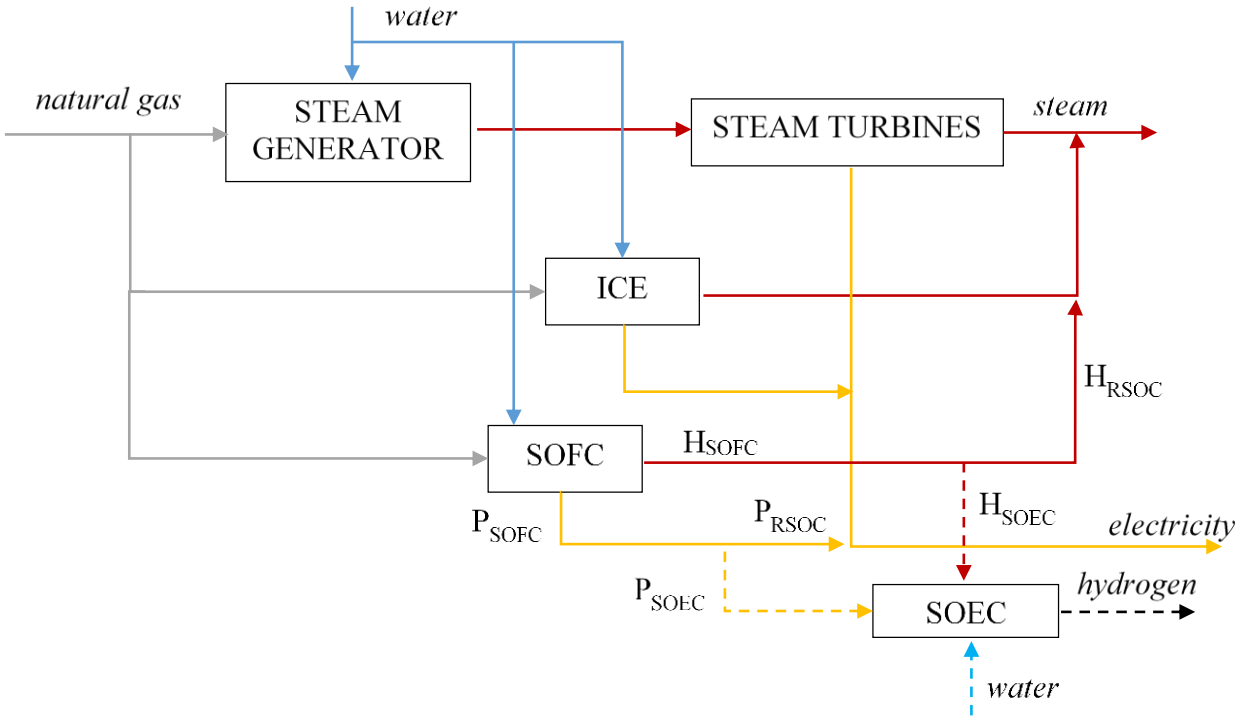


Figure 37. Overview of the novel system. Grey flow is natural gas, blue flow is water, red flow is steam, yellow lines are electricity, and black lines are hydrogen. Inlet water of SOEC is an independent circuit with respect to water for steam production

In the paper mill, Line 1 and Line 2 have different energy demands. The SOFC-SOEC system proposed here features some SOFCs that only operate for combined heat and power generation, whereas other stacks are reversible solid oxide cell. RSOCs are proposed to efficiently manage energy generation: they can be used to produce electricity and steam when both production lines are in operation, whereas they can produce hydrogen (working as SOEC) when only one line works and the energy request is lower. The focus is on performance and efficiency improvements at system level, so the analysis mostly looks at overall values. The term 'RSOC' refers here to the difference between SOFC generation and SOEC consumption, i.e. the net effect of the set of the cells, as seen from the industrial facility. Table 12 resumes RSOC parameters on SOEC and SOFC mode.

RSOC Mode	Parameter	Value
SOEC	Stack nominal electric power	100 kW
	Hydrogen flow (at 90 % of nominal electric power)	2.63 kg/h
	Heat request from SOEC (at 90 % of nominal electric power)	15 kW
	Working temperature	750 °C
	Working pressure	7.01 bar
SOFC	Stack nominal electric power	100 kW
	Electric efficiency (at 90 % of nominal electric power)	52.7 %
	H/P ratio (at 90% of nominal electric power)	0.842

RSOC Mode	Parameter	Value
	Exhaust gases temperature	1330 °C

Table 12. RSOC system parameters.

1.6.6 OPERATION OF THE NOVEL SYSTEM

As previously mentioned, the novel energy generation system proposed is based on three different components: steam turbine, ICE and RSOC. The main goals of the proposed revamping of the cogeneration plant are:

- to increase the efficiency of the energy generation system;
- to cast off the old and low-efficiency steam turbines;
- to produce hydrogen by SOECs during periods of mismatched heat-to-power ratio between the user request ($(H/P)_{USER}$) and the possible SOFC-only generation ($(H/P)_{SOFC}$), achieving the equality by varying the ratio between RSOC working as SOFC and RSOC working as SOEC.

Paper mill could work also in three different mode: only line 1 (case 1), only line 2 (case 2) and both lines (case 1+2). Each case has different energy demands of electricity and heat, consequently different configurations of the novel system are required. In order to size each component, the case with the highest energy load (Case 1+2) is considered. Some constraints are considered:

A. the current generation system will be upgraded gradually, and not entirely dismissed. SOFCs are thought to be adopted to decrease the amount of electricity produced by the oldest component, so with low efficiency, of the system (steam turbine);
B. the existing ICE has better performance than the steam turbines, and it currently works with a base-load strategy. Steam turbines, instead, work as an additional system to cover heat demand. A similar strategy is proposed for the novel system: ICE will work as base load, while SOFCs, and possibly one of the two steam turbines, will work as additional generation system.

The following assumptions are also considered:

- RSOC is used only in SOFC mode due to high electricity consumption;
- only Turbine 1 is used in order to dismiss the highest-power one (Turbine 2), constraint A;
- ICE, steam turbine, and RSOC cover heat and electricity consumption without any integration from grid;
- ICE electricity production is maximised (constraint B).

A system of equations (Eq. 46) describes energy flows and interactions between steam turbine, ICE and RSOC. Electricity (P_{USER}) and heat (H_{USER}) demands from the user are covered by steam turbines cogeneration system (respectively P_{ST} and H_{ST}), ICE (P_{ICE} and H_{ICE}), and the part of RSOC working in SOFC-only mode (P_{SOFC} and H_{SOFC}). Heat-to-power ratio of each component of the system (steam turbine, ICE and SOFC) is defined as constraint, imposing the lower power limits of ICE, steam turbine and SOFC. The steam flow rate varies between 5 ton/h and 21 ton/h, as mentioned in Table 10. Among the solutions of the system, one is selected by solving as an optimisation problem: maximise electricity production of ICE (P_{ICE}) (constraint B).

$$max(P_{ICE})\ s.t.\begin{cases} P_{ICE}+P_{ST}+P_{SOFC}=P_{USER} \\ H_{ICE}+H_{ST}+H_{SOFC}=H_{USER} \\ H_{ICE}-P_{ICE}*f_{ICE,\frac{H}{P}}=0 \\ H_{ST}-P_{ST}*f_{ST,\frac{H}{P}}=0 \\ H_{SOFC}-P_{SOFC}*f_{SOFC,\frac{H}{P}}=0 \\ 5\ <\ \dot{m}_{steam}\ <\ 21 \\ P_{ICE}>0 \\ P_{ST}>0 \\ P_{SOFC}>0 \end{cases} \quad \text{Eq. 46}$$

As previously mentioned, Case 1+2 has the highest energy load because both Line 1 and Line 2 work simultaneously. Since this is the scenario with the highest energy load, it is assumed that P_{SOFC} proposed as solution of this system is, also, the maximum electric power of the RSOC system installed ($P_{RSOC\ installed}$). In other cases, RSOC system can work partly in SOFC mode, partly in SOEC mode. A constraint is added: the sum of electric power of the cells working as SOFC and as SOEC must be lower or equal than $P_{RSOC\ installed}$.

A new system of equations (Eq. 47) describes the novel system when Line 1 or Line 2 works separately (respectively Case 1 and Case 2). In these cases, less energy (both electricity and heat) is required, so a part of RSOCs is used as SOECs to produce hydrogen. The energy request of SOEC (P_{SOEC} and H_{SOEC} for electricity and heat respectively), the equation of heat-to-power ratio of SOEC, and constraint on electric power of SOFC and SOEC (RSOC system) are considered in Eq. 47. Similarly to Eq. 46, maximisation of ICE electricity production (P_{ICE}) is imposed.

$$max(P_{ICE})\ s.t.\begin{cases} P_{ICE}+P_{ST}+P_{SOFC}=P_{USER}+P_{SOEC} \\ H_{ICE}+H_{ST}+H_{SOFC}=H_{USER}+H_{SOEC} \\ H_{ICE}-P_{ICE}*f_{ICE,\frac{H}{P}}=0 \\ H_{ST}-P_{ST}*f_{ST,\frac{H}{P}}=0 \\ H_{SOFC}-P_{SOFC}*f_{SOFC,\frac{H}{P}}=0 \\ H_{SOEC}-P_{SOEC}*f_{SOEC,\frac{H}{P}}=0 \\ 5\ <\ \dot{m}_{steam}\ <\ 21 \\ P_{ICE}>0 \\ P_{ST}>0 \\ P_{SOFC}+-P_{SOEC}\le P_{RSOC\ installed} \end{cases} \quad \text{Eq. 47}$$

Results of Eq. 46 and Eq. 47 of the novel system are reported in Table 13. It is possible to cast off turbine 2 (the highest power one) because the availability of steam from SOFC decreases the steam flow rate to be processed by steam turbine. Adoption of RSOC system allows matching *H/P* ratio of energy production and *H/P* of energy demand. RSOC system produces hydrogen when only

one line operates. Simulations show that electricity consumption from the grid is avoided, making the system grid independent.

CASES	CURRENT SYSTEM			INNOVATIVE SYSTEM			
	ICE (% nominal load)	*Operating turbine*	*Turbine steam flow rate*	*ICE (% nominal load)*	*Turbine steam flow rate*	*SOFC*	*SOEC*
CASE 1	100%	Turbine 1	12.8 ton/h	60%	11.38 ton/h	2.448 MW	576 kW
CASE 2	100%	Turbine 1	8.8 ton/h	100%	6.04 ton/h	2.595 MW	551 kW
CASE 1 + 2	100%	Turbine 2	24.8 ton/h	100%	20.68 ton/h	3.259 MW	-

Table 13. Working parameter of ICE (% partial load), steam turbine (steam flow), SOFC and SOEC electric power.

1.6.7 PRIMARY ENERGY SAVING

This section presents a thermodynamic analysis comparing the current and the novel systems. The analysis calculates energy flows and estimates primary energy (*PE*) consumption for each case. Primary energy consumption is proposed as a benchmark: for each case (Case 1, Case 2 and Case 1+2), primary energy saving (*PES*) between the current and the novel systems is determined (Eq. 48). Primary energy of the existing system ($PE_{current\ sys.}$) is a function of natural gas consumption by steam turbines ($Fuel_{cons,ST}$) and by ICE ($Fuel_{cons,ICE}$), and electricity from grid (E_{grid}). Primary energy of the novel system ($PE_{novel\ sys.}$) is a function of natural gas consumption by steam turbine ($Fuel_{cons,ST}$), by ICE ($Fuel_{cons,ICE}$), and by SOFC ($Fuel_{cons,SOFC}$). Primary energy of natural gas is expressed by LHV (49.2 MJ/kg), whereas electricity consumptions are converted to primary energy by no-renewable primary factor $f_{p,nren}$=1.95 according to Italian standard DM 26/06/2015 (corresponding to an efficiency η_{grid} equal to 0.513). If hydrogen is produced, its primary energy is expressed by LHV (120 MJ/kg), and it is subtracted from the energy consumption of the energy system because it is considered to be used as fuel within the system.

$$PES = 1 - \frac{PE_{novel\ sys.}}{PE_{current\ sys.}}$$

$$PES = 1 - \frac{Fuel_{cons,ST} + Fuel_{cons,ICE} + Fuel_{cons,SOFC} - Fuel_{H2}}{Fuel_{cons,ST} + Fuel_{cons,ICE} + \frac{E_{grid}}{\eta_{grid}}} \qquad \text{Eq. 48}$$

Calculations show that the novel system presents a primary energy saving in all cases. Table 14 summarizes the results: Case 1 and Case 2 have lower *PES*, while hydrogen is produced. Instead,

Case 1+2 does not have hydrogen production (RSOC is used only in SOFC mode due to the higher energy request), and it has the highest efficiency. In fact, SOFC and *PES* (considering also Table 13) are related: the higher SOFC nominal power, the higher *PES*.

CASE	HYDROGEN PRODUCTION	PES
Case 1	16.857 kg/h	2.7%
Case 2	16.137 kg/h	2.3%
Case 1+2	-	6.5%

Table 14. Primary energy saving values for each analysed case.

1.6.8 HYDROGEN GENERATION PERFORMANCE

Results of section 1.6.7 demonstrates that primary energy savings can occur with respect to current system, whereas the complexity of the novel system with hydrogen production is increased. In this section, focusing on hydrogen only, an estimation of the primary energy saving obtained with the production of hydrogen within the novel system with respect to a more established alternative that uses PEM electrolysers (PEMEC) is proposed.

In general, SOEC systems have higher efficiency on hydrogen production compared to other technologies (e.g., PEM electrolysis), thanks to the thermodynamically more favourable operating conditions. However, they require heat at high temperature, which might not be easily available.

In the paper mill application, when the industrial facility operates at partial load (i.e., only one paper production line works) the energy generation system operation also changes: a part of the RSOC units works as SOFC, and a part works as SOEC, thus determining a net hydrogen output. Looking at Figure 38, SOEC uses energy (both electricity and heat) generated only by some of the units in the RSOC system that work as SOFC. SOEC requires heat at high temperature (750 °C, Table 8) which SOFC could provide by means of heat recovery while generating electricity. As SOEC has a proportionally lower heat consumption than electricity compared to the *H/P* ratio of SOFC generation (Table 8), a part of the SOFC heat generation (H_{SOFC}) is used for SOEC (H_{SOEC}), and a part is used to produce steam (H_{SOFC}-H_{SOEC}). Figure 38 details the energy flows for the hydrogen-generating section of the SOFC-SOEC system.

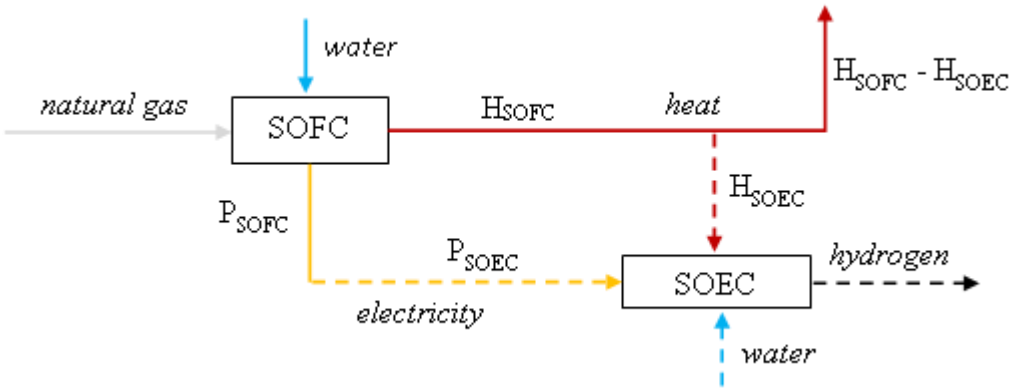


Figure 38. Section of RSOC system used to produce hydrogen. Electricity required by SOEC is provided by SOFC, with no excess (P_{SOFC}=P_{SOEC}); meanwhile a part of the heat (H_{SOFC}-H_{SOEC}) produces steam for industrial needs.

Eq. 49 describes primary energy saving on hydrogen generation ($PES_{H2\ gen}$). The primary energy consumption by the novel system ($PE_{H2\ novel\ sys}$) corresponds to the natural gas consumed by the SOFC when working to produce energy for SOEC ($Fuel_{cons,SOFC,\ H2\ gen}$). The traditional system consumes primary energy ($PE_{H2\ trad\ sys}$) to produce hydrogen by a PEM electrolysis cell, considering that the electricity consumed (E_{PEMEC}) is taken from the grid (η_{grid}, as previous mentioned, is fixed at 0.513). Efficiency on hydrogen production by the PEM electrolyser is fixed at 48 kWh_e/kg_{H2} as proposed in reference [38] for 2020. Finally, to consistently compare the two systems considering the same total output, it is supposed that the traditional system uses a natural gas boiler (efficiency η_{boiler} equal to 0.9) to generate the heat for industrial uses (i.e., the thermal power that is not consumed by SOEC in the novel system, represented by the variable (H_{SOFC}-H_{SOEC}) in Figure 38).

$$PES_{H2\ gen} = 1 - \frac{PE_{H2\ novel\ sys.}}{PE_{H2\ trad\ sys.}} = 1 - \frac{Fuel_{cons,SOFC\ H2\ gen}}{\frac{E_{PEMEC}}{\eta_{grid}} + \frac{H_{SOFC} - H_{SOEC}}{\eta_{boiler}}} \qquad \text{Eq. 49}$$

The comparison only analyses case 1 and case 2 among the paper mill operating configurations, as partial operation of the paper production lines is a necessary condition for net hydrogen production to occur. Eq. 49 supplies the primary energy savings, which are reported in Table 15. In Case 1 and Case 2, values of $PES_{H2\ gen}$ show around 45 % reduction of primary energy consumption with the use of the integrated SOFC-SOEC system. Case 1+2, instead, has no hydrogen production, because RSOC is used only in SOFC mode.

CASE	H_2 PROD	Fuel cons., SOFC	E_{PEMEC}	H_{SOFC}-H_{SOEC}	PES H_2 gen
Case 1	16.857 kg/h	1.09 MW	1.86 MW	0.39 MW	45.6%
Case 2	16.137 kg/h	1.05 MW	1.78 MW	0.36 MW	45.3%

CASE	H_2 PROD	Fuel cons., SOFC	E_{PEMEC}	H_{SOFC}-H_{SOEC}	PES H_2 gen
Case 1+2	-	-	-	-	-

Table 15. Primary energy saving on hydrogen generation

1.6.9 HYDROGEN COST

The aim of the system revamping is not only to save primary energy for the industrial heat and power consumption, but also to introduce a hydrogen production that could be profitable at a reasonable cost. To evaluate the economics of the hydrogen production in the novel system, the RSOC integrated solution is compared to a simpler alternative based on PEMEC. It is considered, also, that when a RSOC system is used, products are both hydrogen and heat for industrial use (Figure 38). Hence, the PEMEC system must be coupled to a steam generator, which is assumed to be a natural gas-fired boiler with efficiency equal to 0.90, to offer the same output. The analysis assumes to feed the PEMEC with grid electricity.

The cost analysis first calculates the equivalent annual cost (*EAC*), then divides it by annual hydrogen production to obtain the fuel cost. *EAC* (Eq. 50) depends on Net Present Value (*NPV*) and Annuity factor ($A_{t,i}$):

$$EAC = \frac{NPV}{A_{t,i}} \qquad \text{Eq. 50}$$

The *NPV* here considered includes investment costs (purchase costs of the components), annual energy costs, and annual maintenance costs (considering inflation rate). Purchase and maintenance costs of RSOC and PEMEC are estimated according to [38] referring to 2020 scenario. RSOC is still a developing technology, so the cost is a forecast and it is likely subject to variation. A sensitivity analysis on the RSOC purchase cost is performed in order to take into account this uncertainty, comprising a range between –10 % and +30 % of the proposed cost. It is considered, also, that the RSOC system integrated in the paper mill produces hydrogen only during 2/3 of the total time, while for 1/3 its components are used in SOFC-only mode (Table 13). Therefore, investment cost is proportionally allocated to the output, and only 2/3 of it affect the hydrogen economics. Variable cost for energy input in terms of electricity and gas are given by [64] and [65], respectively. Moreover, each of them has an annual increasing index, calculated as the average cost variation over the last 4 years. As previously mentioned, heat is available and used to produce steam in the RSOC system, while the PEMEC system uses a natural gas-fired boiler to generate the steam: in this case, only natural gas consumption is considered. Annuity factor is defined by the expected lifetime of the system (assumed equal to 10 years) and by an interest rate of 4 %.

PARAMETER	VALUE
Purchase cost RSOC	2 000 €/kW

PARAMETER	VALUE
Purchase cost PEM	1 000 €/kW
Maintenance cost RSOC/PEM	5 %
Natural gas cost	0.0237 €/kWh
Natural gas increasing index	-0.031 %
Electricity cost	0.1436 €/kWh
Electricity increasing index	-0.059 %
Expected lifetime RSOC/PEM	10 year
Interest rate	4 %
Inflation rate	2 %

Table 16. Economic parameter of the analysis

Figure 39 represents the results in terms of hydrogen cost as a function of the variation of RSOC purchase cost with respect to the forecasted value. In the whole range considered, the RSOC system has a lower cost of hydrogen production than the PEMEC system. Hydrogen cost varies between 6 €/kg and 8 €/kg, while it is equal to 10 €/kg with a PEMEC. This reflects the high efficiency of RSOC in both SOEC and SOFC operation, and the availability of heat from the RSOC system energy balances that is used to generate steam for the industrial facility. On the opposite, the PEMEC system has a separate unit for steam generation, thus excluding any synergy. Moreover, the lower electricity-to-hydrogen efficiency is affected by the high price of electricity compared to natural gas.

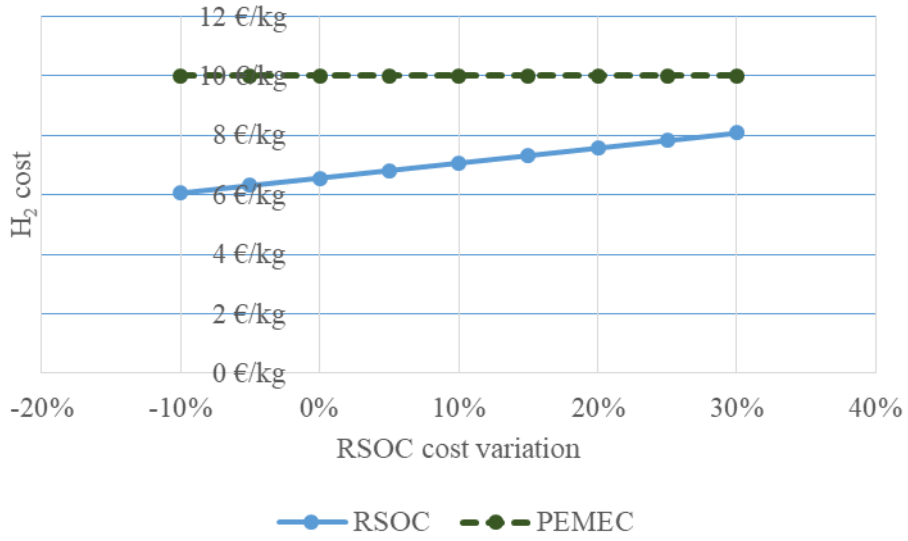


Figure 39. Economic analysis of hydrogen generation cost.

1.6.10 CONCLUSIONS

An innovative energy system based on RSOC for combined heat and power generation has been studied. It is analysed its application to an industrial facility (paper mill) to increase energy efficiency, while adding new business stream in the hydrogen field. The main component of the retrofit is the Reversible Solid Oxide Fuel Cell (RSOC), which can be used alternatively either as SOFC (to generate heat and electricity) or as SOEC (to produce hydrogen). The ratio between SOFC and SOEC varies with the variation of electricity and heat consumption in order to match the H/P ratios at supply and demand side.

The current system is composed by two steam turbines and an ICE, and it allows electricity withdrawn from the power grid when needed. Operational data of the system has been used to model the existing components and size the units of the innovative system. The introduction of RSOCs makes it possible to dismiss one of the low-efficient steam turbines, thus increasing the efficiency on energy generation. Simulations show that it is possible to achieve a primary energy saving up to 6 %: the higher the SOFC power, the higher the achieved PES. Hydrogen is produced at a rate of 16 kg/h, but it occurs exclusively when only one paper production line is in operation. Furthermore, simulation results highlight that the RSOC system features a primary energy saving on hydrogen production in the order of 45 % with respect to a traditional system based on PEM electrolysis fed with grid electricity. Economic analysis has investigated hydrogen generation cost in the proposed RSOC system, comparing it to the production via PEM electrolysers. In the whole range of variation of the investment cost (-10 % to +30% of the value proposed in literature), the RSOC integrated system has lower cost for hydrogen generation.

2 BIG DATA FOR ENERGY EFFICIENCY

2.1 BIG DATA – INTRODUCTION

2.1.1 INTRODUCTION TO BIG DATA AND MACHINE LEARNING

In the recent years, terms as "big data", "data analytics", "machine learning" have become more and more common, and Information Technology (IT) has increased its importance due to many reasons. As computational and memory costs are decreasing slightly, computational power is increasing with very high rates. As a matter of fact, the amount of data available for analysis are increasing enormously, meanwhile their cost is decreasing. According to Mc Kinsey, the era of the so-called "big data" has arrived [66]. "Big data" can be briefly described as 4V: *Volume, Velocity, Voracity* and *Value* [67]. The first V, *Volume*, refers to the amount of data available: they are not just some sample of experiments, but GB or TB of data collected and stored. *Velocity* refers to the frequency of data acquisition, updating and processing: it can be lower than 10 minutes because data needs to be processed and analysed just after they are collected. *Variety* refers to the different sources and structures data can be collected from, for example measurement sensors, weather data and social media. *Value* means that data and (in particular) its analytics increase their value. Reference [68] analyses why big data can impact also on the energy sector: as sensors increase the amount of data available, wireless transmission, network communication and cloud computing technologies increase the transmission and processing velocity. More recently, available sources (also of unstructured energy data) have increased too: not only weather and social media data, but also grid equipment, asset management data (for example generator, transformer etc), smart meter, economic data and Geographic Information System (GIS) are available (what previously was defined as variety).

Not only big data, but also machine learning methods can be used in the energy sector. Machine learning was defined as "*the field of study that gives computers the ability to learn without being explicitly programmed*" (Arthur Samuel - 1959) or similarly "*A computer program is said to learn from experience E with respect to some class of tasks T and performance P, if its performance at tasks in T, as measured by P, improves with experience E*" (Tom Mitchell - 1997). These methods can directly learn the task to do from dataset even if the process is unknown and/or it is impossible to be described with equations. Different methods have already been defined and are available ([69]); some of them, that are strictly related to energy, are briefly summarized in the next paragraphs.

2.1.2 CLUSTERING AND KNN

Clustering methods such as k-means are used to divide dataset into homogeneous groups. At least one distance function must be defined to calculate the distance between each datum of the dataset. The main purpose of this method is the classification task. It is labelled as *unsupervised method* because it learns on its own how to classify data. Figure 40, Figure 41 and Table 17 represent an example [70] of clustering applied to Iris dataset. Iris dataset contains 50 sampling data (sepal

width and length, petal width and length) of 3 different species of Iris, *Setosa*, *Versicolor* and *Virginica*. Data are classified using k-means algorithm by variables sepal width and length, petal width and length: the algorithm divides data into homogeneous groups using a distance function (for example Euclidean distance) calculated on the values of these variables. Each group (cluster) aggregates samplings with the lowest distance between each other. If three clusters are chosen, it is expected that k-means divides sampling data into the three species previously cited. It is possible to appreciate that clustering allows to correctly classify mostly of the dataset with the correct species (144 on 150 of total observations);

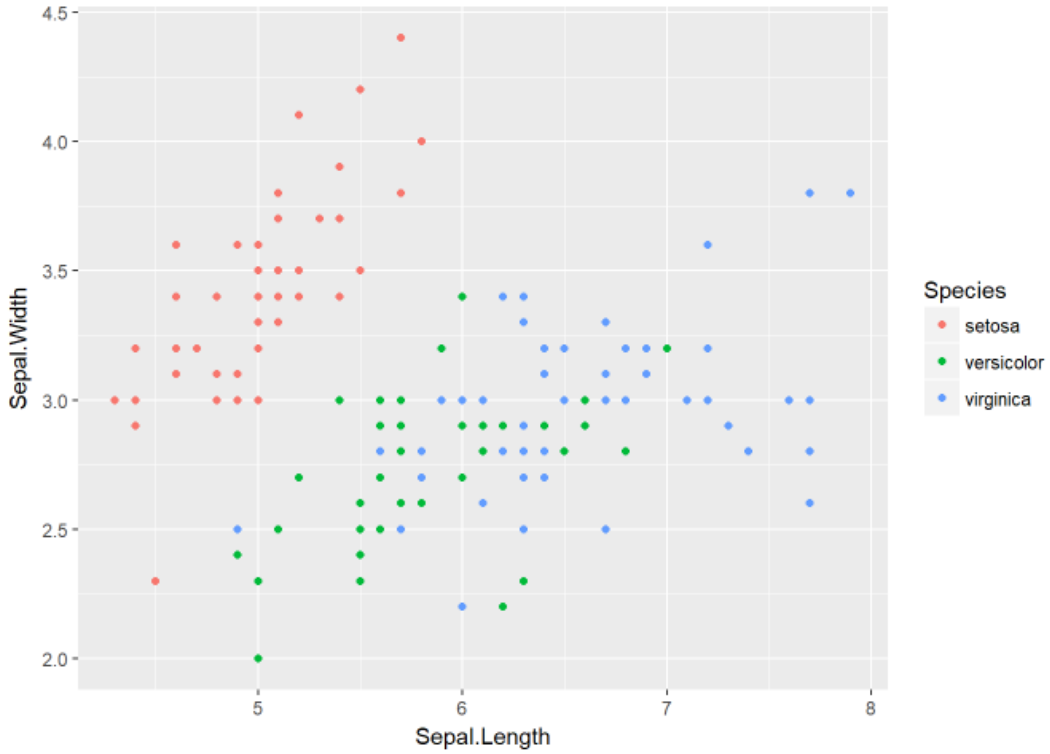


Figure 40. Iris dataset plotted over sepal width and sepal length [70]

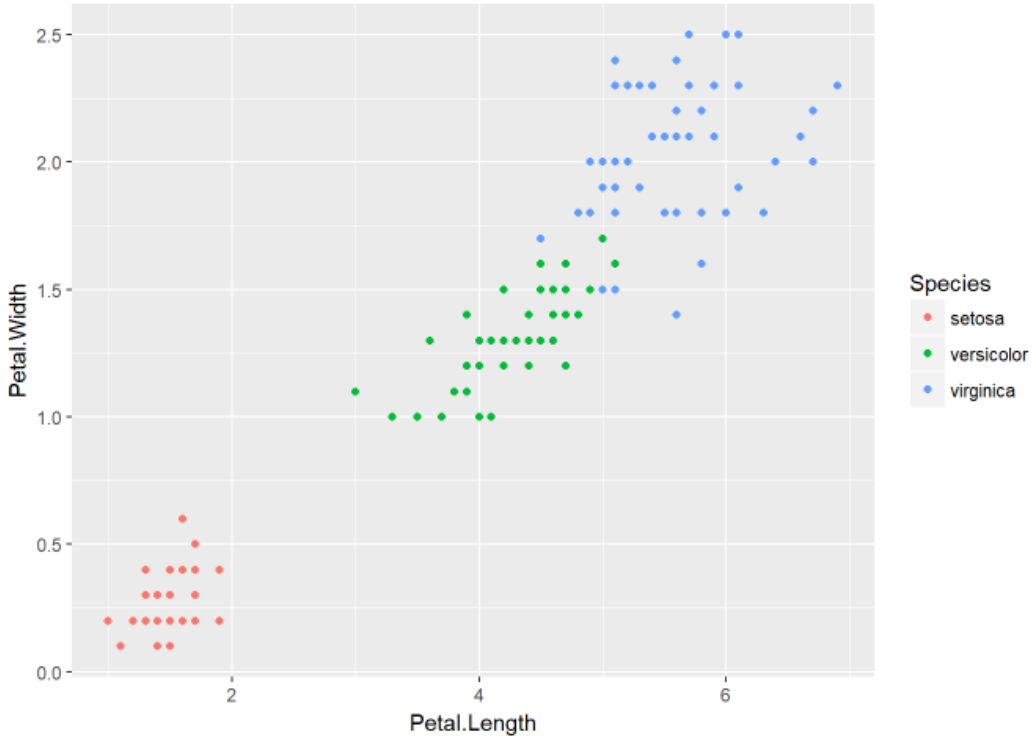


Figure 41. Iris dataset plotted over petal width and petal length [70]

		Dataset		
		Setosa	**Versicolor**	**Virginica**
Clustering	**Setosa**	**50**		
	Versicolor		**48**	4
	Virginica		2	**46**

Table 17. Iris dataset classification [70]

k-Nearest Neighbour (kNN) is similar to the clustering methods previously cited. It is used mainly for classification purposes. It is a "supervised method", that means that classification is performed using a dataset previously classified.

2.1.3 PRINCIPAL COMPONENT ANALYSIS

Principal Component Analysis (or also PCA) is probably the most well-known method of *dimensional reduction* [71]. The aim of this method is to decrease the dimensions of the dataset analysis by using an orthogonal transformation. Consequently, also the variety of data decreases. It is expected that most of the phenomena observed have linear correlate variables even if this correlation is unknown. When PCA is performed, resulting variables will be uncorrelated. Data analysis and analytics are easier when PCA is applied because a lower number of variables must be plotted. Reference [72] is an example of a dimensional reduction problem: the author uses PCA to

decrease the number of variables of a thermal energy system (solar absorption chiller system) to predict the system performance (Figure 42).

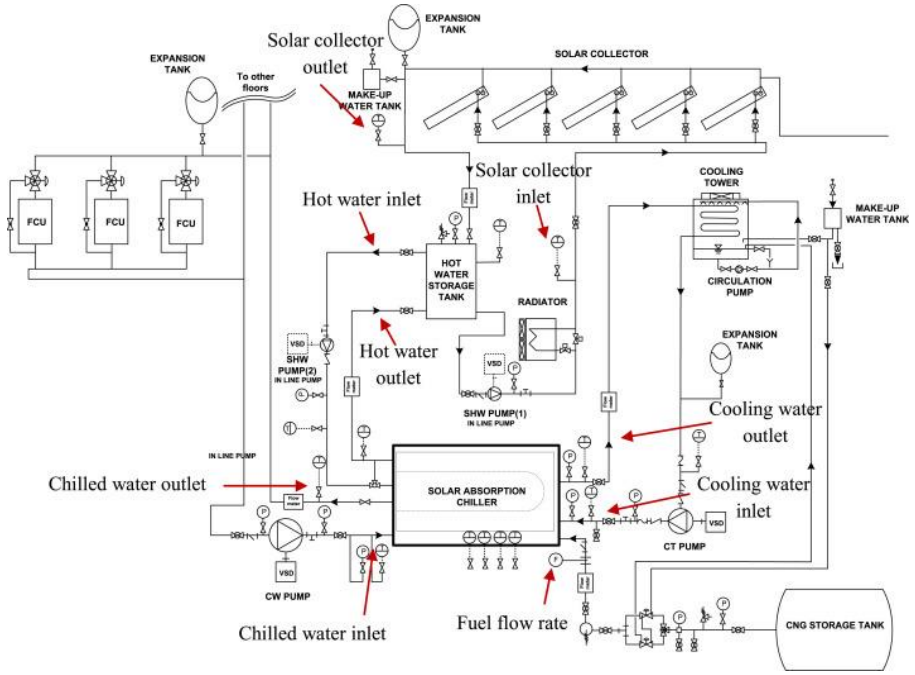


Figure 42. Thermal energy system [72]

When PCA is applied, it is possible to analyse which is the contribution of each variables to PCA components (Table 18). Value can vary from -1 to 1 as R parameter of the linear regression: the closer to 1, the variable and PCA component are closely related.

Variable	**PCA Component**			
	PCA 1	**PCA 2**	**PCA 3**	**PCA 4**
$T_{chw,i}$	0.0530	0.1185	−0.0910	−0.0857
$T_{chw,o}$	−0.0180	0.1539	−0.1152	−0.0146
$T_{hw,i}$	0.5102	−0.0862	−0.2571	−0.0857
$T_{hw,o}$	0.3850	0.1626	0.5644	0.1345
$T_{cw,i}$	0.0974	0.2508	0.4163	0.1062
$T_{cw,o}$	0.1540	0.2764	0.3216	0.0597
$T_{SC,i}$	0.4736	0.0207	0.0385	0.0349
$T_{SC,o}$	0.5678	−0.0867	−0.3357	−0.2005
V_{fuel}	−0.0445	0.8736	−0.3883	0.0569
T_{amb}	0.0804	−0.1342	−0.2282	0.9526

Table 18. PCA applied to a thermal energy system for dimensional reduction [72]

It is possible to appreciate from Figure 43 that the first PCA component (PCA 1) resumes hot water inlet, hot water outlet, solar collector outlet and solar collector inlet variables. Such variables give the higher contribution to the first PCA component, and they are strictly correlated each other as they are related to solar collectors (Figure 42). For this component, dimensional reduction occurs

because four variables are described with just one PCA component. Figure 44 represents the contribution to PCA component 2: only fuel flow gives a high contribution to this variable, consequently it means that for this component a dimensional reduction is not applied.

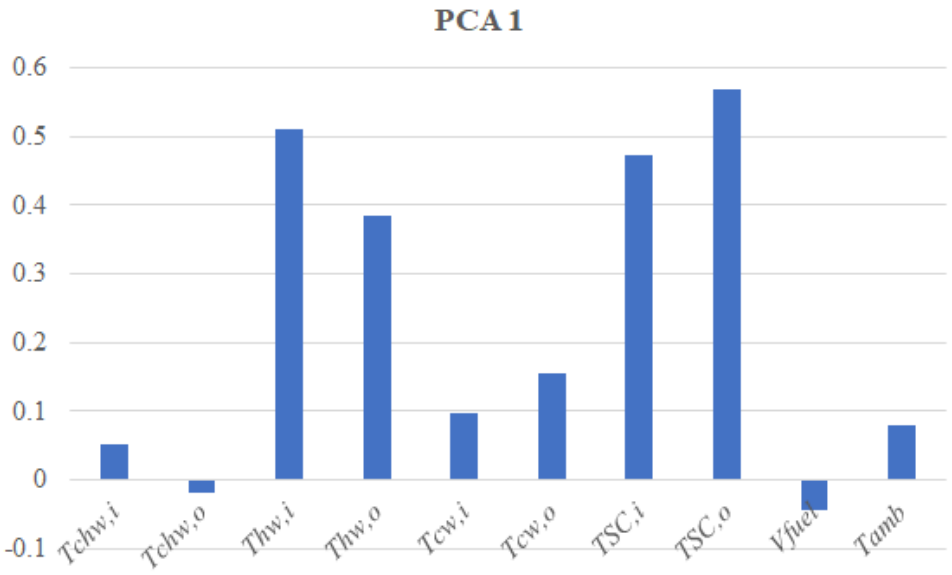


Figure 43. PCA component 1 over variables [72]

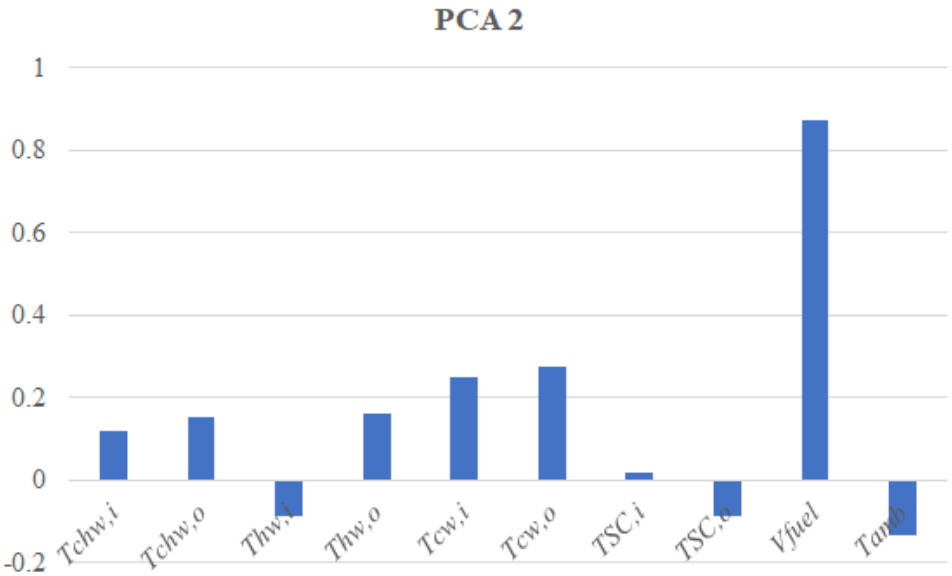


Figure 44. PCA component 2 over variables [72]

2.1.4 DECISION TREE

A decision tree is a decision support tool where the outcomes are based on the values of the algorithms that contain only conditional control statements (Figure 45).

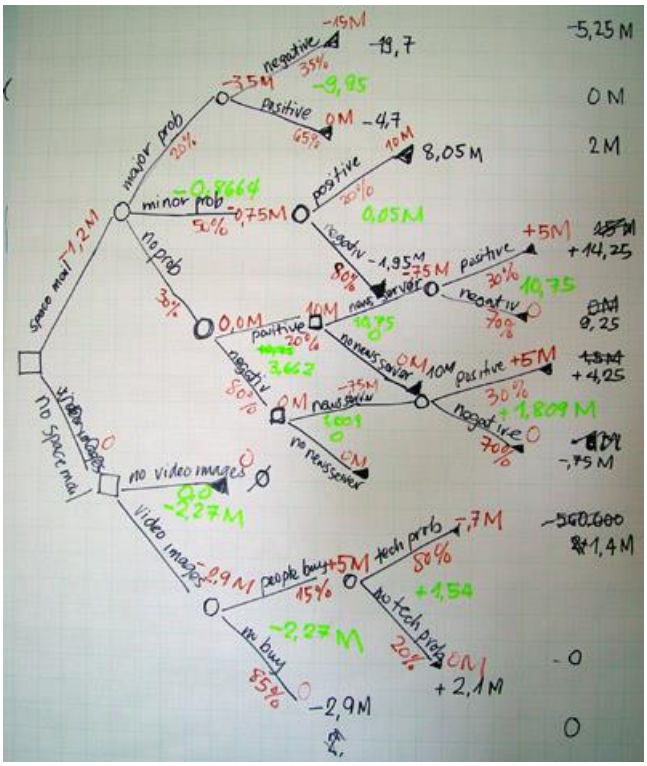


Figure 45. Decision tree example (Image from [73]*)*

Decision tree learning is a machine learning method that uses observations to define decision tree mainly for classification, regression and data mining. Even if the phenomena described in the dataset are unknown, this method learns from observations and defines condition statements to create an algorithm to calculate outcomes by minimising errors. Reference [74] uses decision tree machine learning to predict energy demand of a HVAC system of a hotel by using outdoor air temperature, relative humidity, number of rooms booked and previous value of electricity consumption. Such methodology analyses data and proposes an algorithm based only on conditional statements of the variables: the result is the decision shown in Figure 46 where the predicted consumption depends on each step of the tree.

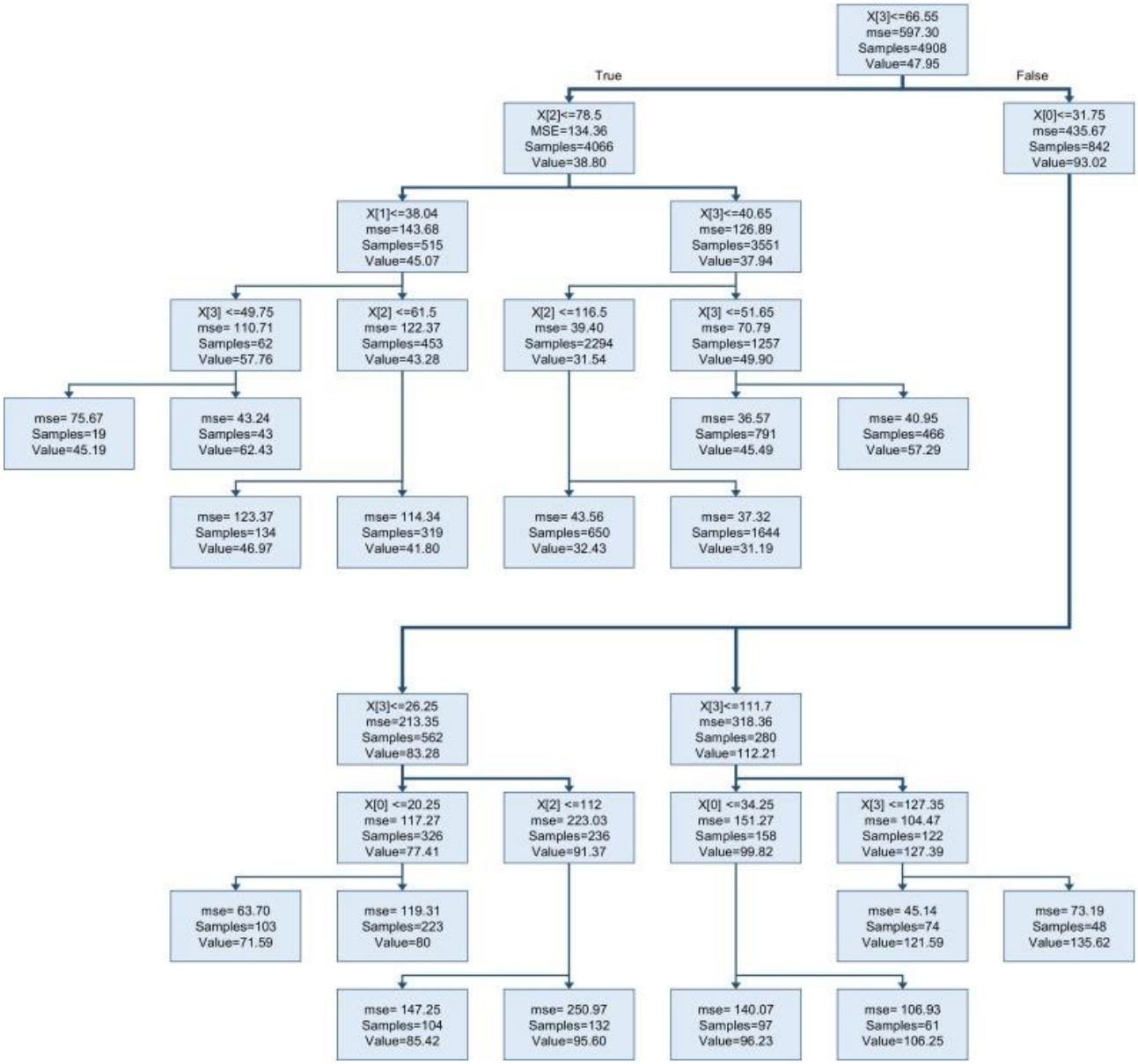


Figure 46. Decision tree example [74] *– X[O] is outdoor air temperature, X[1] is relative humidity, X[2] is number of room booked, X[4] previous value of electricity consumption*

2.1.5 ARTIFICIAL NEURAL NETWORK

Artificial neural network (ANN) is probably one of the most known machine learning methods because it mimics the brain. The brain is made by neurons connected to each other by axons; our experiences modify the connections between neurons, as axons can increase or decrease connections depending on how much they are used. The elementary unit of an ANN is the artificial neuron, which processes the incoming signal to other neurons. Artificial neurons are connected to each other and a weight on the incoming signal is usually defined. ANN must be trained: they do not have any knowledge about the tasks they have to perform, so some examples must be given in order to train the entire network. This method is highly flexible because it could be used to perform tasks such as image recognition, classification and definition of black box. ANN is defined “black box” model because it is defined only by observed value and it is unknown how the model is set up, in particular

which are the connections and the weights between of the neurons. In the previously cited reference [72], an ANN is used to predict the performance ($Q_{cooling}$ and COP) of a solar absorption chiller system. Each variable selected for the model is connected to a neuron of the input layer, whereas the performance variables are connected to the neurons of the output layer. Between the input and the output layer, a hidden layer is added (in this case only one even if more could be used); the connections between the input, hidden and output layers are defined by observed data with a training process (Figure 47). Each neuron has a transfer function (Figure 48) to transform input to output: in this case, tangent sigmoid and linear functions are reported to be used.

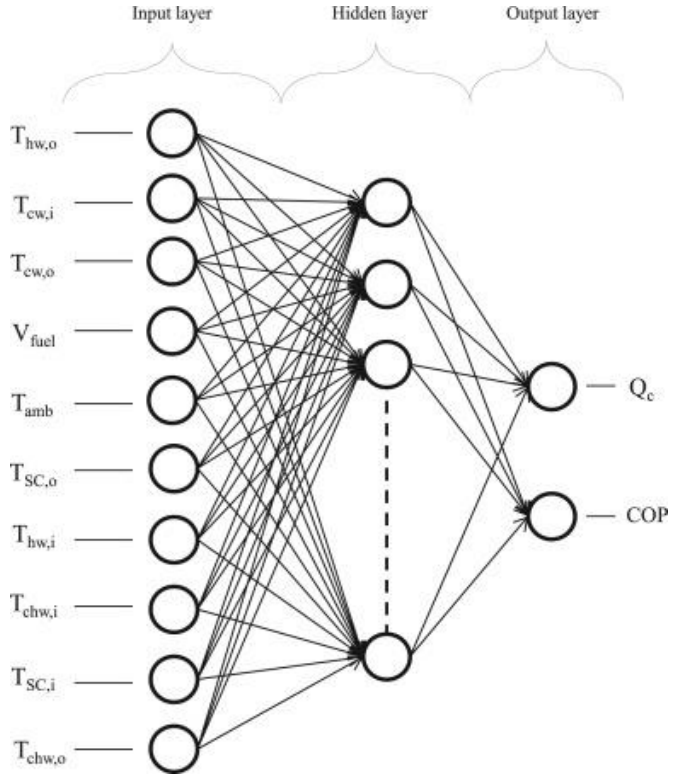


Figure 47. ANN representation [72]

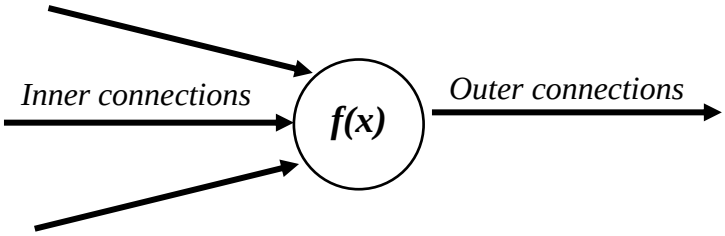
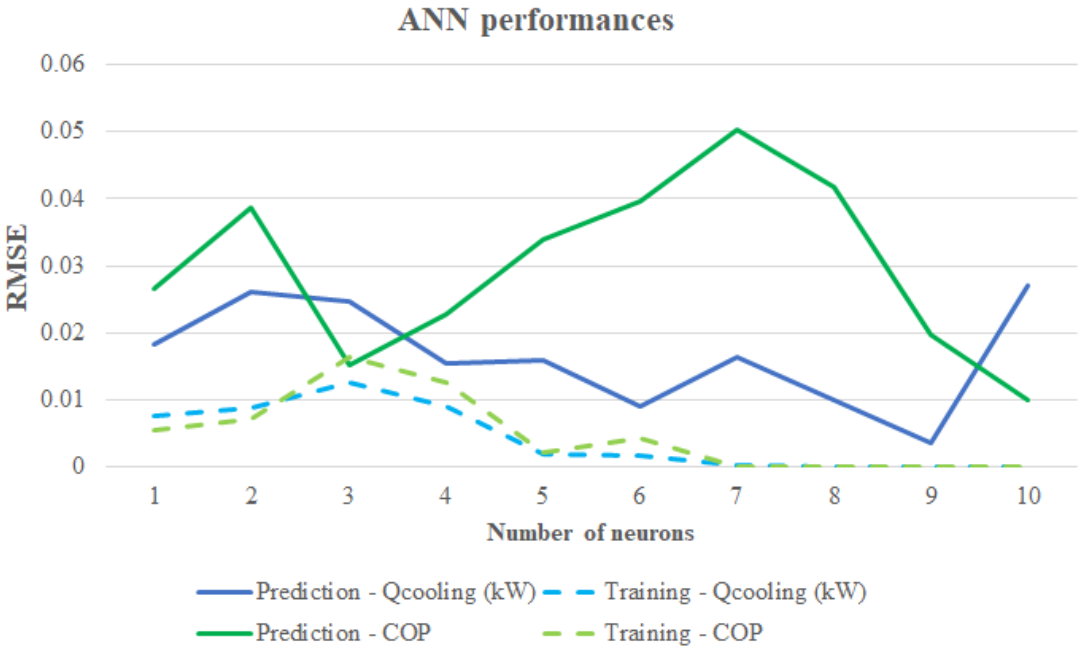


Figure 48. Neuron representation. Inner and outer connection are related to a transfer function f(x)

By varying the number of neurons of the hidden layer, ANN performance varies too. Figure 49 represents the root mean square error (RMSE) on performance ($Q_{cooling}$ and COP): generally, errors

decrease when the number of neurons increases. It is possible to appreciate that 6 neurons on the input layer are sufficient to decrease RMSE.



.

Figure 49. ANN performances on training and prediction varying hidden layer number of neurons (data from [72]*)*

2.1.6 EXAMPLES OF APPLICATION OF BIG DATA AND MACHINE LEARNING IN ENERGY SECTOR

Machine learning methods and big data have been already proposed to be applied to energy sector in many fields. Examples of possible application are:

- *Load forecasting*, machine learning can be useful to analyse time series data and forecast consumption of energy. Ref. [75] proposes ANN with Autoregressive Integrated Moving Average (ARIMA) to analyse time series. References [76]–[78] analyse time series using clustering and weighted kNN algorithm to find similar pattern of consumption of electricity and its price;

- *Smart building*, machine learning methods are proposed to improve the efficiency of energy generation and consumption of buildings. Ref. [79] proposes ANN and genetic algorithms to optimise energy usage and, consequently, reducing consumptions. Ref. [74] previously cited compares ANN and decision tree to forecast HVAC electricity consumption of a hotel. Clustering is proposed in [80] to define patterns of electricity curves of building consumptions;

- *Energy production*, machine learning method can be useful to analyse energy system, define grey/black box of the system, and to estimate the performances and optimisation. Ref. [81] reports that ANN are used to estimate thermal efficiency, specific fuel

consumption and volumetric efficiency under different partial load and methane ratio of internal gas engines using biogas. Ref. [82] proposes ANN to simulate solar collector and estimate its performances at different working condition with the aim of identifying possible research gap and improvements. Ref. [83] simulates an absorption system based on water-ammonia using ANN with different designs, solution fractions and working parameters; the scope is to find the optimum design in order to maximize exergy efficiency and the coefficient of performance. Ref. [84] improves geothermal energy availability analysing data with ANN;

- *Renewable energy forecasting and management*, it is well-known that renewable energy systems (RES) may have high variability on production and availability. Machine learning methods are proposed to increase their efficiency, mainly to forecast production by using, for example, weather data. Ref. [85] proposed ANN and multi-linear regression method to estimate solar radiation with daily meteorological measurements. References [86]–[88] propose reviews on photovoltaics power generation forecasting methods: ANN, clustering, decision tree, support vector machine, support vector regression and ARIMA models are proposed. Not only solar energy but also other RES are investigated: ref. [89] reviews different machine learning methods to forecast wind energy for electricity production.

The previous examples want to briefly show how big data and machine learning methods could be applied to the energy sector. In the next chapters, clustering and kNN methods will be analysed to define innovative methods useful for energy system, in particular for the design and optimisation of energy generation system and to increase accuracy of load forecasting. The aim is to increase efficiency of energy generation system in industrial facilities.

2.2 CLUSTERING – INTRODUCTION AND METHODS

2.2.1 INTRODUCTION TO CLUSTERING AND K-MEANS

Clustering is a type of machine learning method used to classify data and to perform data segmentation. Samples are grouped into subsets or "clusters": in each cluster, objects are more likely to be related to one another than to those assigned to different clusters. Clustering is strictly related to the concept of "degree of similarity" (or "degree of dissimilarity") between the objects being clustered. Similarity is defined by the method, for example distance function. For each group of data (cluster) it is possible to define a centroid. Figure 50 represents an example: a dataset is clustered into three different homogenous groups and centroids are centred into each group. A datum is classified by its distance from the centroid: it is related to the group with the nearest centroid.

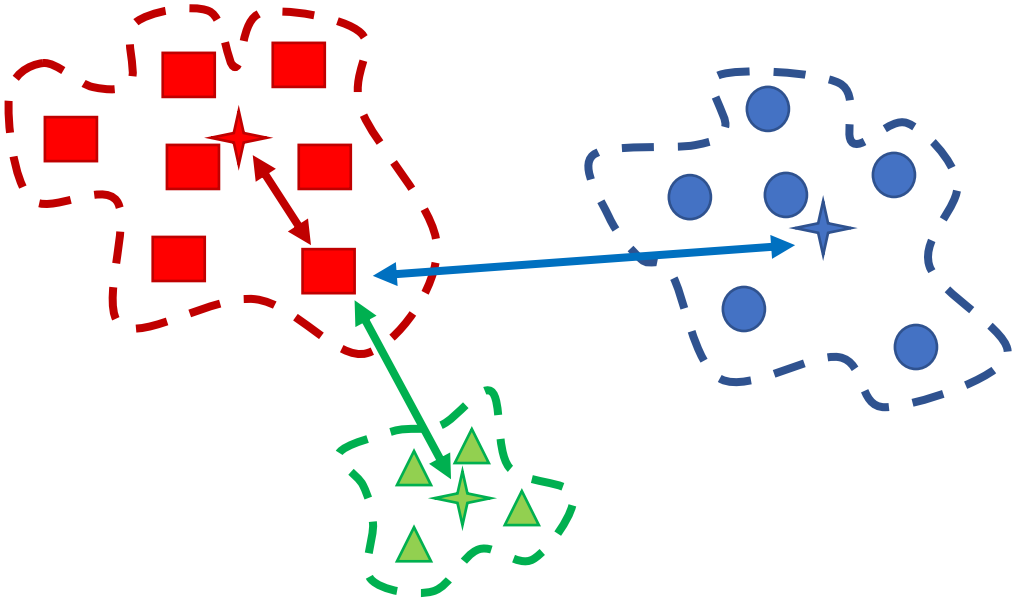


Figure 50. Clustering example

K-means is a clustering method used when all the variables are quantitative and Euclidean distance between objects is defined as a dissimilarity function: the lowest the distance, the highest the similarity. Euclidean distance between each object x_a and x_b is measured using variable i=1...n that describes each object (Eq. 51):

$$d(x_a, x_b) = \sum_{i=1}^{n} (x_a - x_b)^2 \qquad \text{Eq. 51}$$

If a dataset with m objects is provided, K-means divides the dataset into N clusters minimising Euclidean distance between each object of the cluster. The number of clusters N is given by the user. In the next section, the silhouette method is described. It defines the suitable number of clusters for a

dataset. Euclidean distance is not the only distance function for K-means method: for example, MATLAB © ([90]) provides a built-in function (k-means) to perform cluster analysis in which cosine, absolute distance (called also cityblock), correlation and hamming distance functions are defined as well.

2.2.2 HOW TO CHOOSE THE CORRECT NUMBER OF CLUSTERS – SILHOUETTE METHOD

As previously mentioned, cluster analysis requires to define the number of clusters. Silhouette ([91], [92]) method is proposed in order to perform such a task.

Silhouette criterion optimises the number of clusters by searching for the maximum distance between observation of the same cluster and observation assigned to the neighbouring one. Silhouette *s(i)* of the object *i* is defined by Eq. 52:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \qquad \text{Eq. 52}$$

Where:

- *a(i)* is the average distance between the datum *i* and all the other observations assigned to the same cluster;
- *b(i)* is the average distance between the datum *i* and all the observations assigned to the neighbouring clusters.

Average *s(i)* of all the observations is considered. As a matter of fact, varying the number of clusters also average *s(i)* changes: the optimum number of clusters maximises the average silhouette.

Figure 51 represents a dataset referring to the case study described in section 2.4: each point is a sample, and it is described by electric power consumption and heat to power ratio (H/P) of the industrial user. Each observations greater or lower than 6 times of the standard deviation (sigma) are excluded in order to not consider outliers. Silhouette is applied to estimate the optimum number of clusters: it is possible to appreciate that if the number of clusters is 3, the maximum average silhouette can be obtained (Figure 52, also referring to the case study further reported). This condition means that the distance between objects of the same cluster and objects of the neighbouring is maximised.

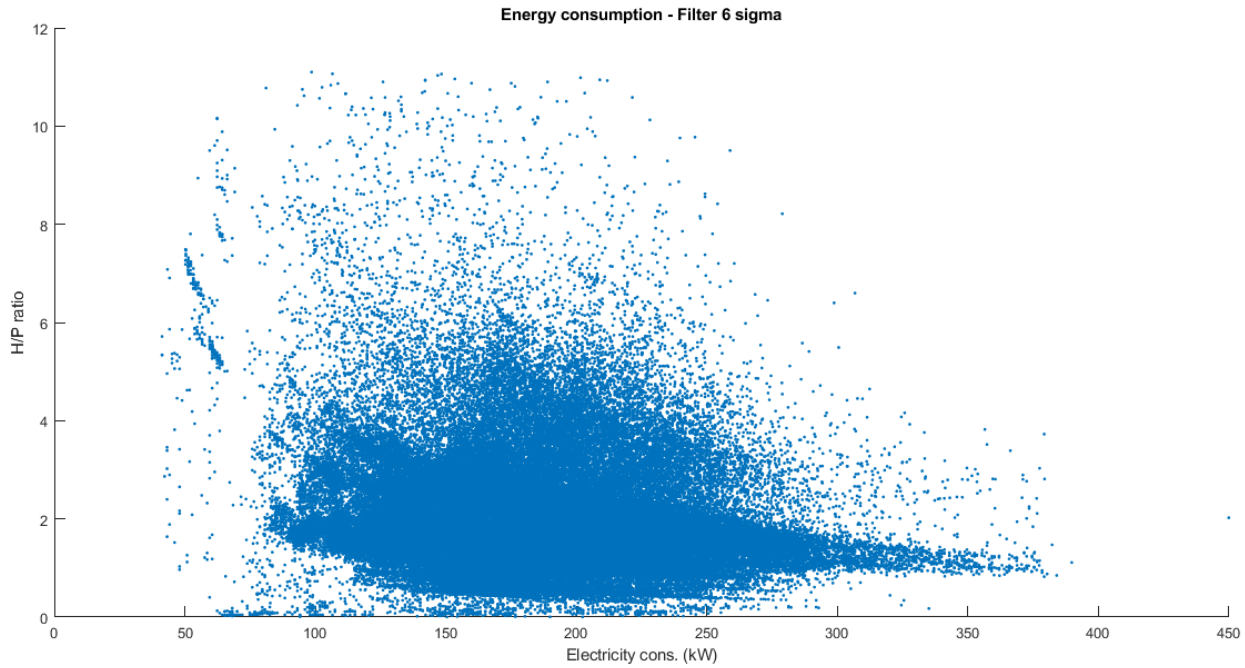


Figure 51. Energy request described by electricity power consumption (kW) and H/P ratio with a 6 times sigma filter applied

Figure 52. Average silhouette of dataset presented in Figure 51

2.2.3 INTRODUCTION TO KNN

kNN (k-Nearest Neighbors) is an unsupervised machine learning method used mainly for classification and regression [69]. If a classified dataset is given, kNN classifies new observations choosing the nearest k neighbors of each observation. Figure 53 represents an example: a new observation (green triangular) must be classified in a dataset where data are classified into red rectangular and blue dot. The observation is classified by firstly calculating the distance with each

datum, and then observing the classification of the k nearest neighbors. If k = 3 is chosen, observation is classified as blue dot category, whereas if k = 5 is chosen, observation is classified as red rectangular category.

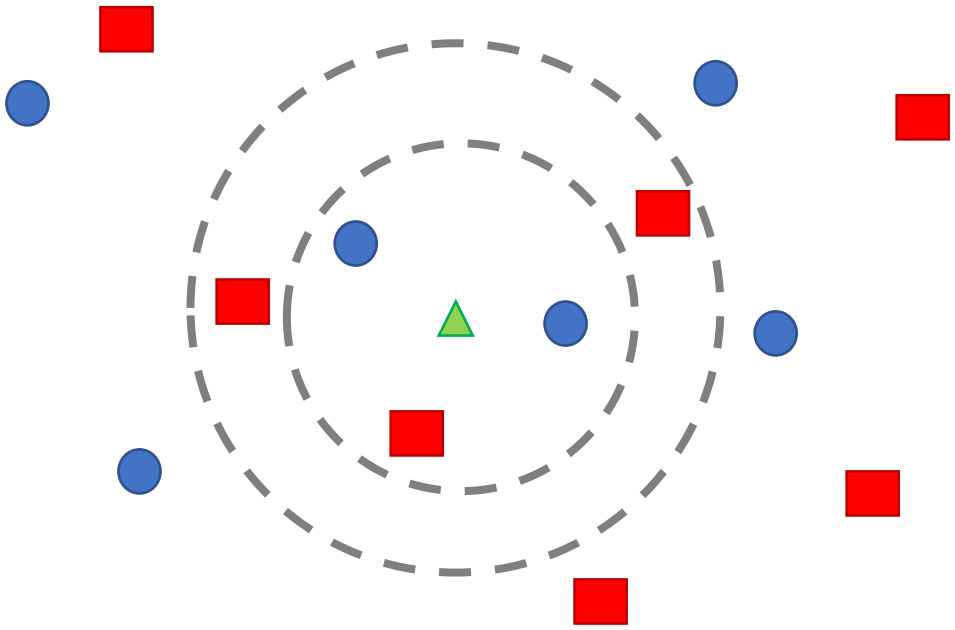


Figure 53. kNN example – Classification of an observation

kNN is useful when a dataset is clustered and then new observations need to be classified. For example, in section 2.5 it is proposed a short forecasting method where a dataset is clustered and then kNN is used for forecast. Clustering is used to divide dataset in N clusters, and for each cluster an average curve is defined. When a new observation occurs, forecast is performed classifying its cluster; consequently, the average curve is proposed. kNN performs the classification task analyzing how the k neighbors nearest to the observation are classified, and the distances between them.

kNN requires two hyperparameters: the number of neighbors (k), and the distance function. These parameters are usually defined by using heuristic techniques or cross-validation. Here we propose to use *FitchkNN* fuction developed in MATLAB ©: it optimizes the kNN model by choosing the distance function and the number of neighbors to decrease the classification error [93].

2.3 USE OF CLUSTERING TO DEFINE POLYGENERATION SYSTEM

2.3.1 INTRODUCTION

As presented in [2], not all the industrial sectors have the same energy intensity, that is energy consumption over production. Moreover, each industrial sector features different ratio between the different types of energy consumed (i.e. different thermal to electricity ratio): drying is more energy intensive compared to assembling. An increase on energy efficiency would decrease the final energy consumption, but each industrial sector requires different energy generation systems and technologies.

Many studies have been already performed with the aim of increasing efficiency. For example, analyses on energy consumption and heat recovery on energy intensive sectors were performed ([55], [56], [94]–[96]).

Polygeneration systems are used when more than one type of energy is requested simultaneously, for example cooling, heat and electricity. More specifically, cogeneration is used when two different types of energy are requested, for example heat and electricity, or cooling and electricity. Different energy systems can be used to cover user's energy request and to achieve high thermodynamic/economic efficiency.

Typically, a cogeneration system is sized on a cumulative curve, representing the number of hours each value of power is requested for (Figure 54). The size of the cogeneration system is chosen in order to cover thermal or electric load for a defined number of hours. The main problem of these graphs is that they analyse separately each type of energy (e.g. electricity and heat), and they do not analyse the relationship between them or common daily pattern of consumption. In the author's opinion, the size of a cogeneration / polygeneration system is a classification problem, because it is necessary to classify energy data to choose the correct system in terms of both technology and size.

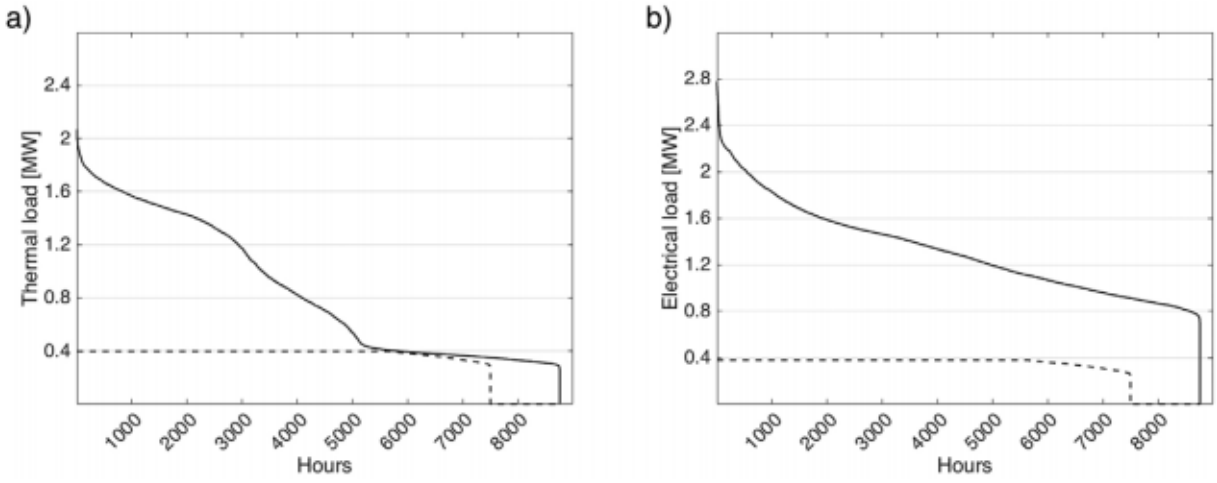


Figure 54. Cumulative curve of thermal (a) and electric load (b), continuous lines represent the energy request, dashed lines the cogeneration production (from [97]*)*

An innovative approach to design energy generation systems based on big data analysis is here developed. More specifically, a study on how cluster analysis could be applied to analyse energy

demand data is depicted. The aim of the method is to design cogeneration systems because they suit the energy demand profiles, by choosing the correct type of cogeneration technology, operation strategy and energy storages. Clustering can improve how a cogeneration system is chosen. As a matter of fact, by dividing data in uniform datasets of energy demand and energy ratio (like H/P), it is possible to choose the most suitable technology for each cluster, thus improving energy performance.

Clustering can be used also to classify energy demand profiles to define standard patterns of consumption. This is useful to evaluate the viability of energy storages in case of mismatching between energy production and consumption and to suggest the most suitable operation strategy. Firstly, it is supposed that the user's energy request data are sampled uniformly (for example every 15 minutes, every hour etc) and stored using at least three variables: time stamp (date and time of observation) and at least two energy variables, for example electric power and mean heat power request. Then data will be used to perform two different analyses, power analysis and profile analysis:

1. **Power analysis**: every observation is considered separately to define clusters with similar values of the variables (i.e. electricity demand and H/P ratio). This information, and how such variables vary inside the cluster, will suggest the most suitable polygeneration technology and/or information to design the generation system;
2. **Profile analysis**: daily energy demand profile (not a single observation) is defined and clustered to identify how energy demand varies during daytime. Possible mismatching can be detected between energy demand and energy production using energy system defined with Power analysis.

In the next paragraph, the proposed methodology will be explained.

2.3.2 METHODOLOGY

Power and Profile analysis are based on clustering of data but with a different definition of the dataset. In the first one, each observation is a datum; in the second one, each datum is a day of observation. A common workflow is proposed (Figure 55) to define a dataset and cluster data, and to perform them:

1. **Data cleaning**: not all the observed data are suitable for the analysis due to measurement errors and/or bad electric signal. Filter should be applied to the dataset to delete outlayers and/or uselnss observations.
2. **Dataset creation**: observed and filtered data are used to create suitable datasets for the analysis. In case of power analysis, each observed data can be a record of the dataset. In case of profile analysis, observations lasting one day can be arranged together to define a daily consumption profile. After a dataset is created, data are normalized.
3. **Estimation of the number of clusters**: before applying clustering, it is necessary to define the number of clusters. Silhouette is applied to estimate the suitable number of clusters for the dataset.
4. **Cluster analysis**: cluster analysis is applied to the dataset.

Figure 55. Analysis workflow proposed

Power analysis is performed with the definition of a dataset where each observation is reported and represented at least by two energy variables, for example electric power and H/P ratio. It is expected that cluster analysis defines homogenous clusters according to the variables. In the case of electric power and range of H/P ratio, it suggests which type of cogenerator technology fits better the data by using properly references such as Table 1 based on [3]. If a significant mismatching between energy generation and demand occurs, it could be suitable to add components to the polygeneration system to best fit the production and demand curves meanwhile increasing overall efficiency of the system. For example, if it is necessary to adjust the H/P ratio of a cogeneration system, heat integration system and/or heat pumps could be used to increase it, meanwhile Stirling engines and ORC would be useful to decrease it by converting unused heat into electricity. Figure 56 represents an example of Power Analysis, where dataset has two energy variables: electricity demand and H/P ratio. Clustering is applied, and dataset divides data into homogenous groups. For each cluster (group), it is possible to appreciate how variables varies: in this case, electricity over H/P ratio. An average H/P curve is proposed to analyse the variation. It is possible to appreciate that electricity demand varies between 80 and 180 kW, meanwhile H/P is around 2. This information can be used to choose the most suitable cogeneration technology (Table 1).

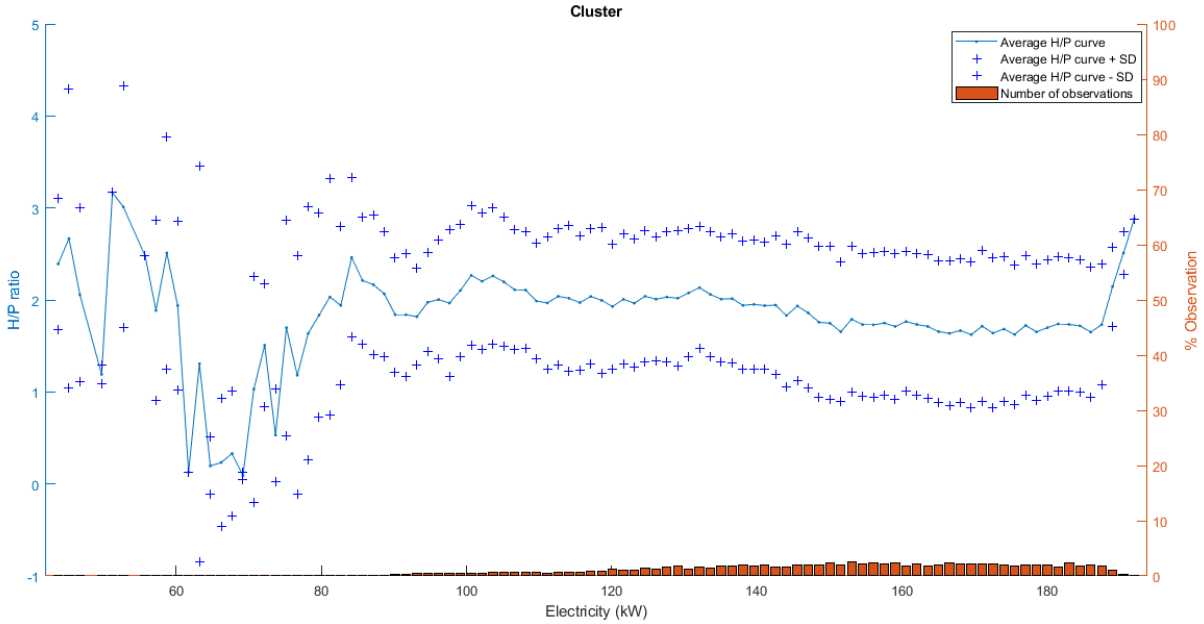


Figure 56. Example of Power analysis – representation of a cluster

Profile analysis is proposed to define different consumption profiles. Each profile could be used to verify how energy demands vary during daytime, that is the best operation strategy to increase efficiency. In order to perform the analysis, it is proposed to create a dataset where each datum is a daily consumption profile; it is described by several variables that depend on the number of observations and on how many energy variables are observed. In the case that energy demand is sampled every 15 minutes with two variables (electric power and H/P ratio), each daily profile is made by 96 observations, and each datum is described by 192 variables (96 variables for the electric power and 96 for the H/P ratio). As a matter of fact, cluster analysis divides the dataset into homogenous clusters; for each cluster, energy demand profiles are similar, and it is possible to define reference curves. Reference curves could be used to analyse the mismatching between energy production by the cogeneration system and energy request. With respect to Power analysis, Profile analysis applied with reference curve gives information on:

- **Energy storage**: if a mismatching between energy production by the cogeneration system and reference profile occurs, energy can be stored when production is higher than demand. The reference curve provides also information on the size of the storage required;
- **Energy integration system**: as previously mentioned, an integration system can occur if the mismatching between energy production and demand cannot be stored;
- **Operation strategy**: reference curves provide useful information concerning which operation strategy for the system is more suitable to increase its efficiency and to decrease the operative costs.

Figure 57 represents a sample of the profile analysis: each datum of the dataset is a day of observation. If clustering is applied, it is possible to define similar profile of consumption. Profile analysis lets understand how each energy variable (in this case electricity consumption and H/P) varies during daytime, and how they are related to each other.

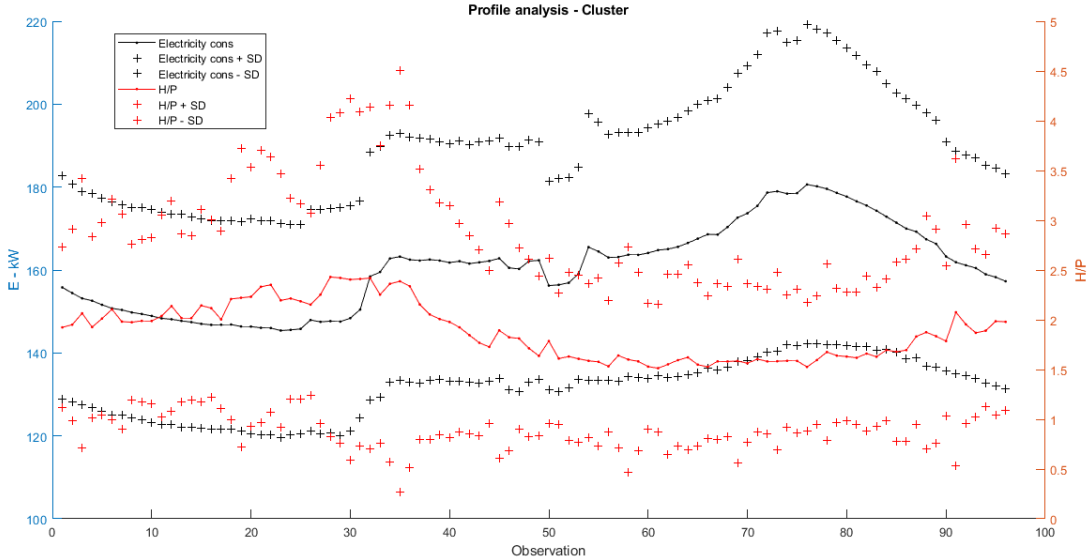


Figure 57. Example of Profile analysis – representation of a cluster

2.4 CLUSTER FOR ENERGY SYSTEM - A CASE STUDY

2.4.1 CASE STUDY INTRODUCTION

A case study is presented to apply the proposed analysis: it concerns an industrial facility selling wood (timber) laminated window, plywood, engineered veneer, laminate, flooring and white wood. The firm is located in the north of Italy with suppliers in East Europe, Africa and a subsidiary in Gabon. Every year it works about 110 000 m^3 of trunks and its ovens could dry 6 000 m^3 of wood per cycle. It was founded more than 100 years ago and the total revenue in 2017 was of over 90 millions of euros. The industrial process requires to dry the wood into kilns and to store it into warehouses. Electricity is used for the production equipment, offices, lighting purpose into the warehouses and to charge electric forklifts. Heat is used for the kilns that work at about 70 °C. Energy is actually generated by using:

- two cogeneration systems (CHP) based on internal combustion engines (ICE) to produce both electricity and heat;
- natural gas fired boiler as an integration system for the kilns when the cogeneration systems do not produce enough heat compared to the one requested.

Electricity can be withdrawn from the grid if the demand is higher that the production. Figure 58 represents the energy fluxes and the interconnections between each component of the system. The proposal is to classify energy demand data with suitable variables in order to determine which energy generation system and operation strategy is more suitable for this application.

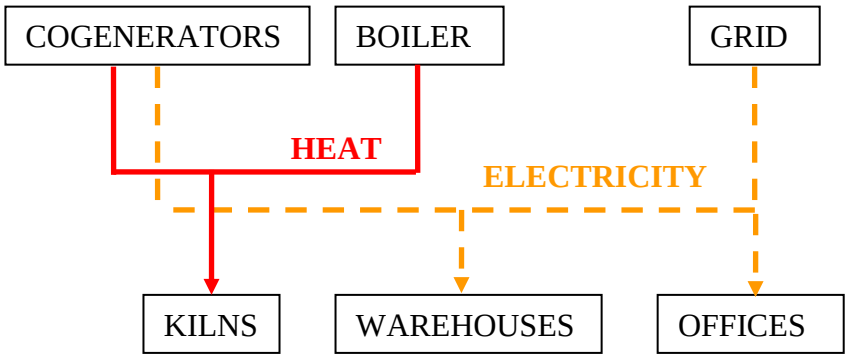


Figure 58. Electricity and heat energy fluxes, connection between production and demand

Energy demand (both electricity and heat) was sampled each 15 minutes from 01/01/2015 to 25/09/2017. Electricity data is available as the mean power requested (kW). The heat request, instead, is calculated by measuring water flow rate (m^3/h) and inlet and outlet temperature (°C) to heat the kilns. Data are stored into a structured SQL database. It is expected that this dataset could contain some sampling event with missing measurement and with outliers. Missing measurement in a SQL database are managed with NULL value, so events with at least one variable with a NULL value are

not considered for the study because the system is not able to sample the process, and other variables could be affected by errors. Outliers could occur because data are stored without any validation.

2.4.2 DATA CLEANING

The first step of the proposed workflow for each analysis is data cleaning. Not only missing errors, but also errors on sensors and/or in recording values may occurs. Data are plotted with a histogram (with a log scale on the x axis) and a probability plot of quartiles (QQ plot) to intercept outliers. QQ plot is used to compared dataset distribution to normal distribution: where there is not matching probably, outliers occur. Figure 59 displays how data are distributed: it is possible to appreciate that outliers are present for both electricity and heat demand. Electricity data are mainly between 100 and 1000 kW, while the maximum sampled value is higher than 10^6 kW. The same occurs for the heat demand: in fact, QQ plots show that the current dataset does not follow a standard distribution.

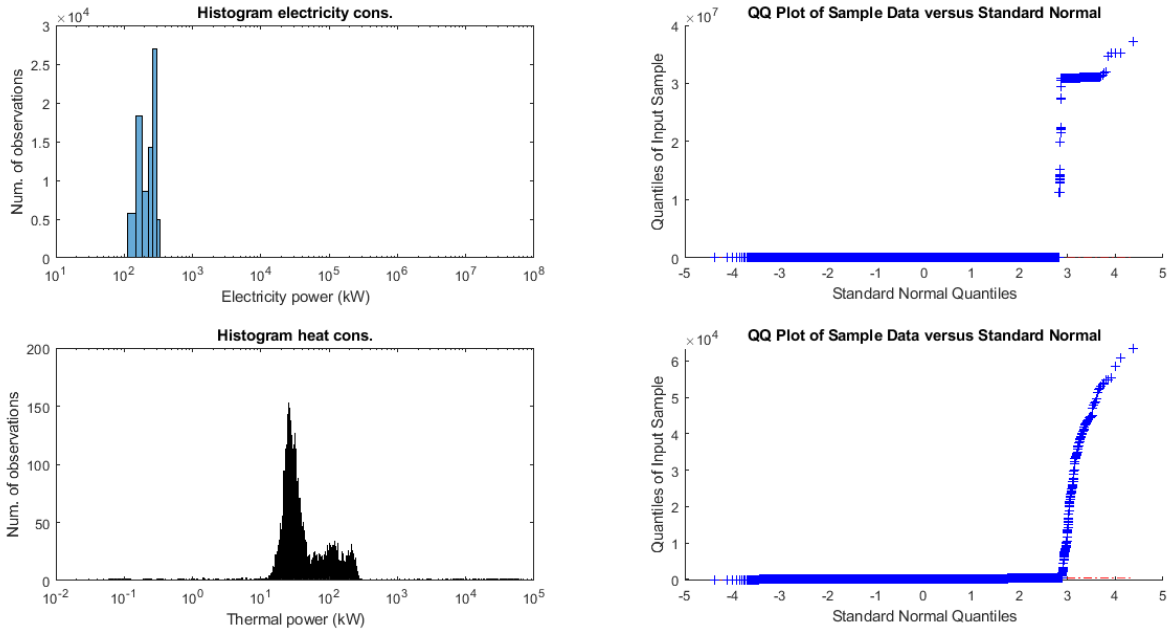


Figure 59. Representation of the dataset without filtering data, histogram and QQ plot of electricity (top) and thermal power (bottom) demand

To filter the outliers, it is proposed to define an upper limit for each of the variables, both for electricity and heat demand. The limit is set considering the maximum request of electricity and heat of the system. Figure 60 represents the filtered data: QQ plots show that the filtered dataset is closer to a normal distribution and the range of the dataset has decreased.

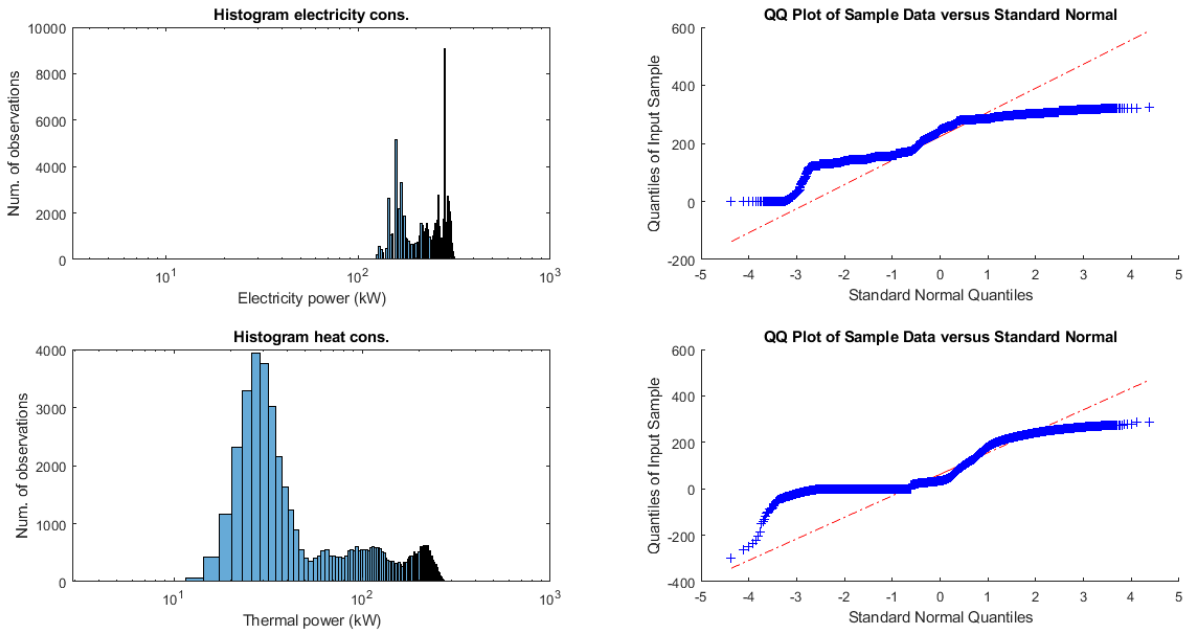


Figure 60. Representation of the dataset filtering data, histogram and QQ plot of electricity (top) and thermal power (bottom) demand

After data cleaning is performed, data can be used to define the dataset for the analysis and to perform them.

2.4.3 POWER ANALYSIS

As previously mentioned, Power analysis is carried out to choose the most suitable energy system. Dataset previously defined and presented in Figure 51 is clustered, where each datum is an observation with two variables, electricity power and H/P ratio. Silhouette criterion is applied to define the most suitable number of clusters (Figure 52). Figure 61 represents the dataset after being clustered into 3 groups, while Table 19 resumes each cluster. Cluster 2 and cluster 3 resume more than 85% of the observations, with a H/P ratio between 0 and 4.72. Electricity power demand varies from 190 kW to 390 kW for cluster 2, and from 43 kW to 192 kW for cluster 3. According to, Table 19 the most suitable cogeneration technologies for such values of electricity power and H/P range are ICE and gas micro-turbines. It is possible to appreciate that the actual cogeneration system is based on ICE. The gas micro-turbines system allows a H/P range higher than ICE technology: a solution with two turbines of 200 kW each is proposed because considering only one turbine is useful to cover the energy request represented in cluster 2, but not in cluster 3.

For each cluster, a reference curve (H/P ratio vs electricity power) is defined based on observations in order to analyse possible mismatching between observations and the proposed energy system. Figure 62, Figure 63 and Figure 64 represent respectively the reference curves of cluster 1, cluster 2 and cluster 3. For each reference curve, it is represented the mean of the observations varying

electricity power and H/P ratio: mean observations curve ± one standard deviation (Mean obs + 1 SD and Mean obs - 1 SD). In each figure, the cogeneration system proposed is simulated by following the electricity demand (cogeneration system curve). For each point of the mean observation curve, it is displayed the number of the observations in the dataset: this variable shows how many times this value of electricity power occurs. It is possible to appreciate for cluster 3 (Figure 64) that consumption occurs more frequently between 120 and 200 kW: the cogeneration system here proposed suits perfectly the reference curve. For cluster 2, instead, simulations show that the cogeneration system has a H/P ratio higher than the requested (Figure 63): it is necessary to evaluate a suitable operation strategy and/or an energy storage. Cluster 1 (Figure 62) has electricity demand between 100 and 220 kW, meanwhile H/P ratio is between 4 and 7: a heat integration system is necessary because H/P request is higher than H/P of the cogeneration system proposed.

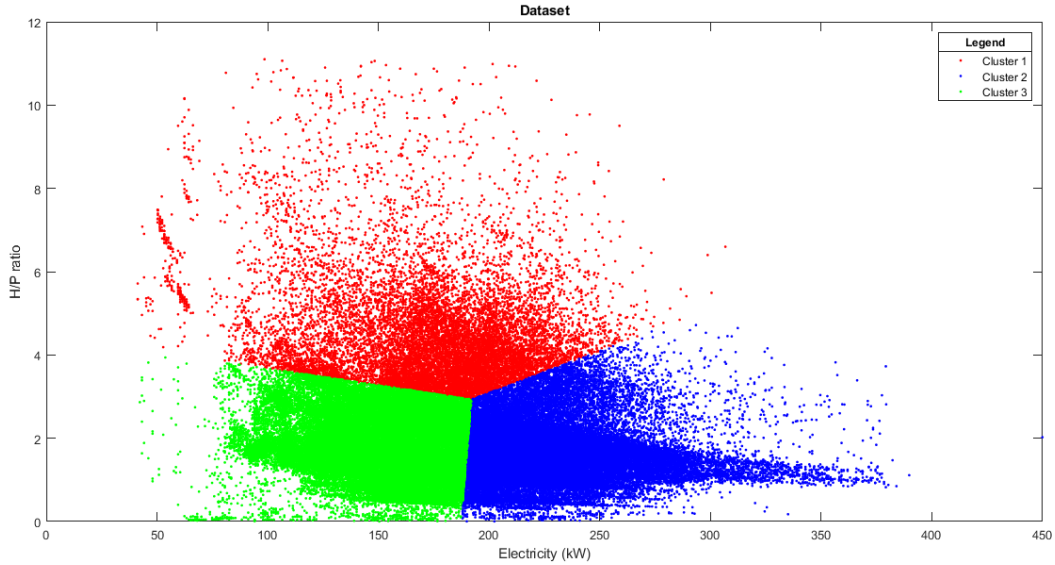


Figure 61. Clustered data of Figure 51 for Analysis 1 – Power Analysis

Table 19. Cluster and number of observations for each cluster of Power analysis

Cluster	Number of observations	Electricity cons. range	H/P ratio range
1	13.34 %	43 – 306 kW	2.9 - 11
2	42.13 %	190 – 390 kW	0 – 4.72
3	44.53 %	43 – 192 kW	0 – 3.9

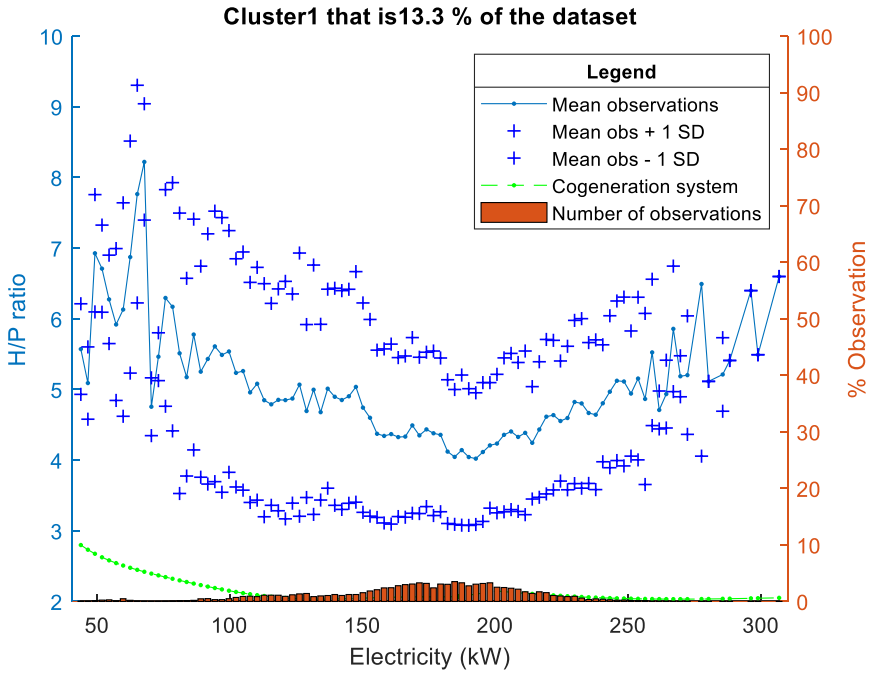


Figure 62. Reference curve of electricity – H/P of cluster 1 (13.3 % of the observations)

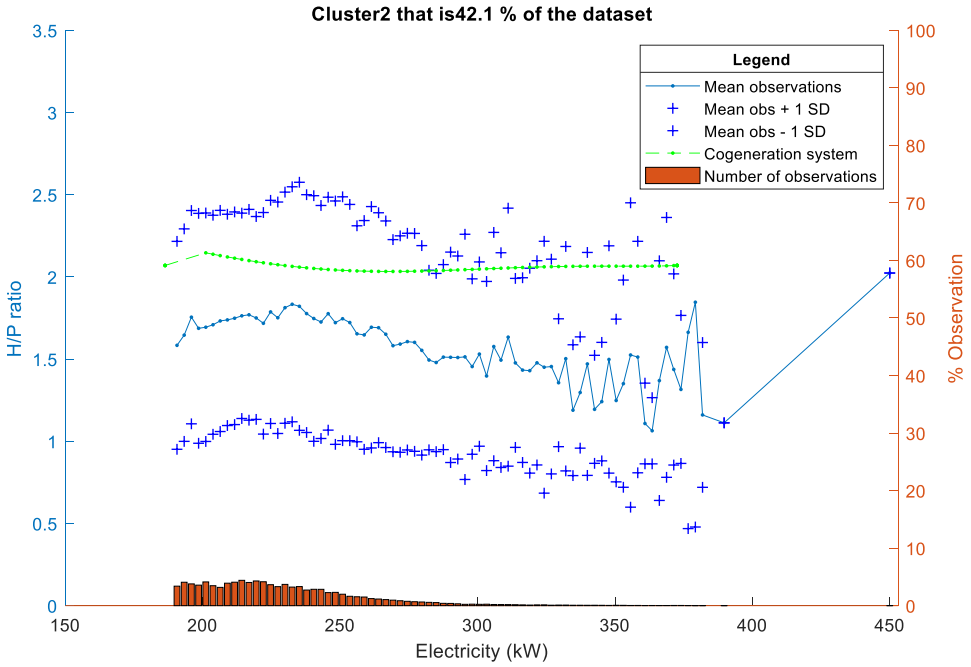


Figure 63. Reference curve of electricity – H/P of cluster 2 (42.1 % of the observations)

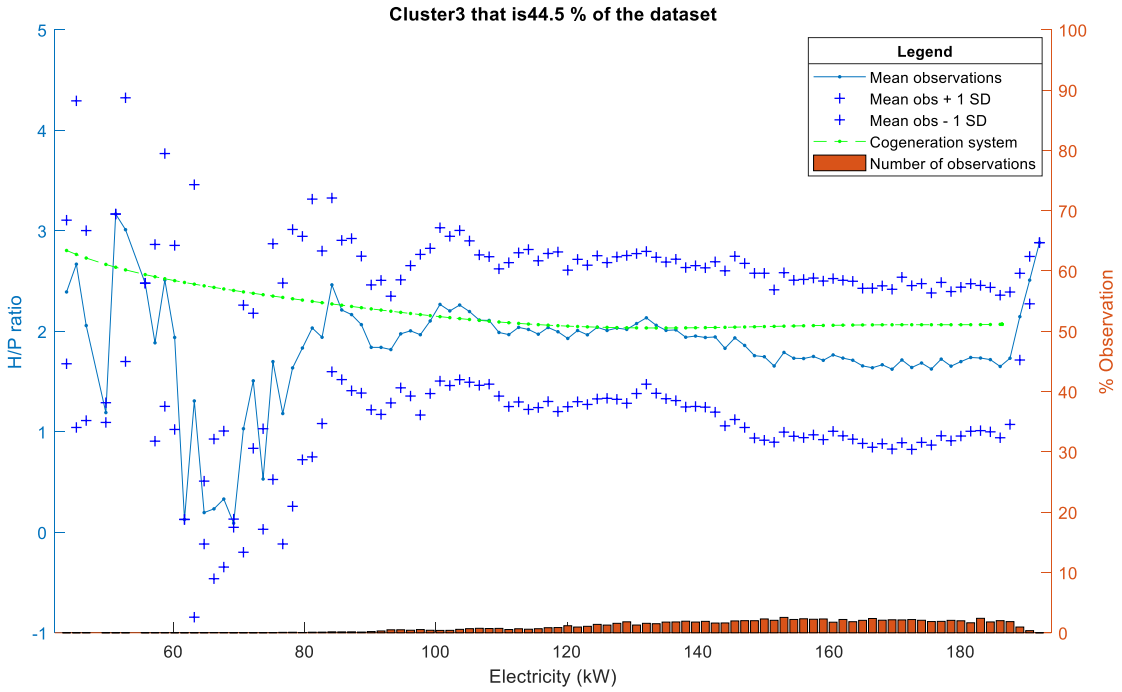


Figure 64. Reference curve of electricity – H/P of cluster 3 (44.5 % of the observations)

2.4.4 PROFILE ANALYSIS

Profile analysis is performed to verify the necessity of energy storage systems, and to check for the most suitable operation strategy. Each datum of the dataset is a daily profile, the cluster analysis is used to check similar patterns of consumption and to define average profiles. In the case study here proposed, the silhouette criterion suggests dividing the dataset into 4 clusters (Figure 65). Even if the maximum distance is 2, there is a local maximum with 4 clusters. It is chosen to increase slightly the number of clusters (from 2 to 4) in order to have average profiles that suit better the observations. The proposed cogeneration system is plotted as well under the assumption that it follows the electricity demand. Table 20 resumes the importance of each cluster: cluster 4 describes more than 45% of the total observations and cluster 1 about 32%. The analysis of the average curves of these two clusters describes more than 75% of the sampled days.

Table 20. Cluster and number of observations for each cluster of Profile Analysis

Cluster	Number of observations
1	31.91 %
2	21.90 %
3	0.27 %
4	45.92 %

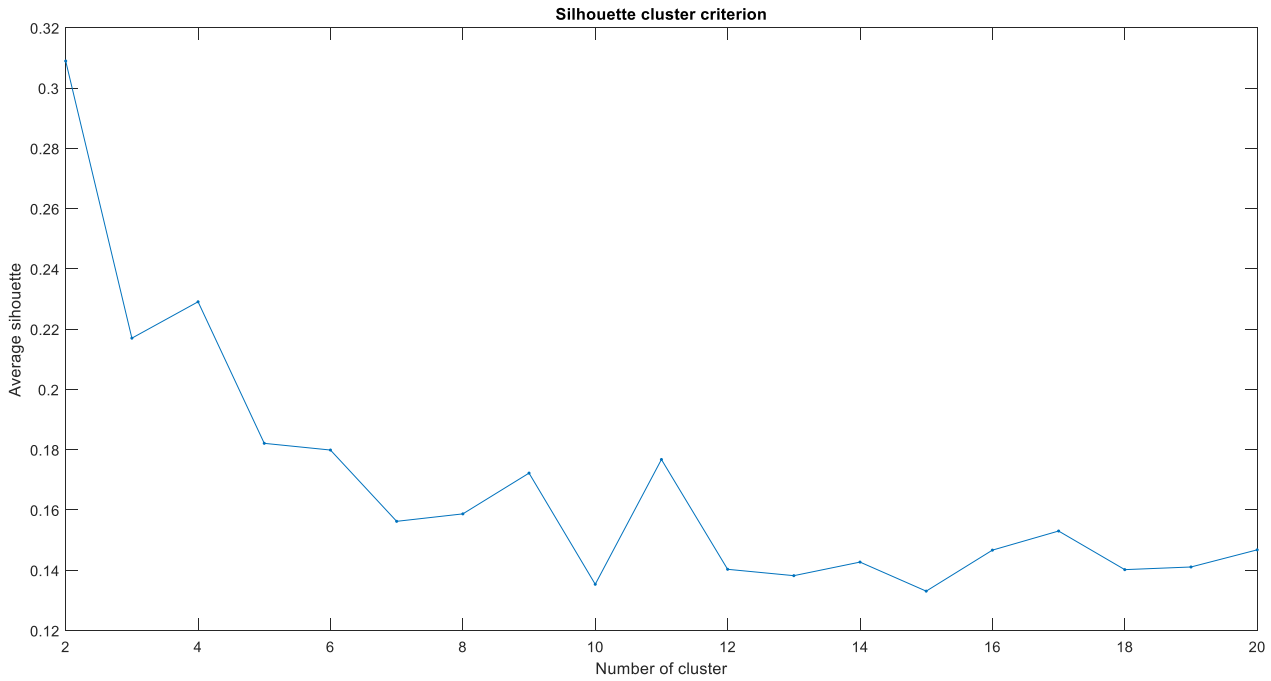


Figure 65. Average silhouette for Analysis 2

Figure 66, Figure 67, Figure 68 and Figure 69 represent the average electricity and H/P ratio profile for cluster 1, cluster 2, cluster 3 and cluster 4 respectively. In the x-axis, it is reported the time of the observation: each profile represents one day of observations, sampled every 15 minutes. The first observation is at 0:00, and the last one (96^{th}) at 23:45. For each figure, three curves are reported. The mean electricity consumption curve shows how it varies during daytime (left y axis). A range of ± one standard deviation is added (Electricity cons. + 1 SD and Electricity cons. - 1 SD). The second curve (Mean H/P cons.) represents the average H/P consumption during daytime (right y axis); a range is defined considering ± one standard deviation (H/P cons. + 1 SD and H/P cons. - 1 SD). H/P cogen. system curve represents the H/P of the cogeneration system here proposed when it follows the electricity consumption (Mean electricity cons.). Mismatching between H/P request (Mean H/P cons.) and H/P of the system (H/P cogen. system curve) may occurs: if Mean H/P cons. is higher, more heat is required compared to the heat available, consequently an integration system is required. On the contrary, if Mean H/P cons. is lower, a quantity of heat is available and unused. A heat storage can be useful in order to increase overall efficiency of the system as heat is not dissipated when it is available, and a heat integration system is used only when the heat storage is empty.

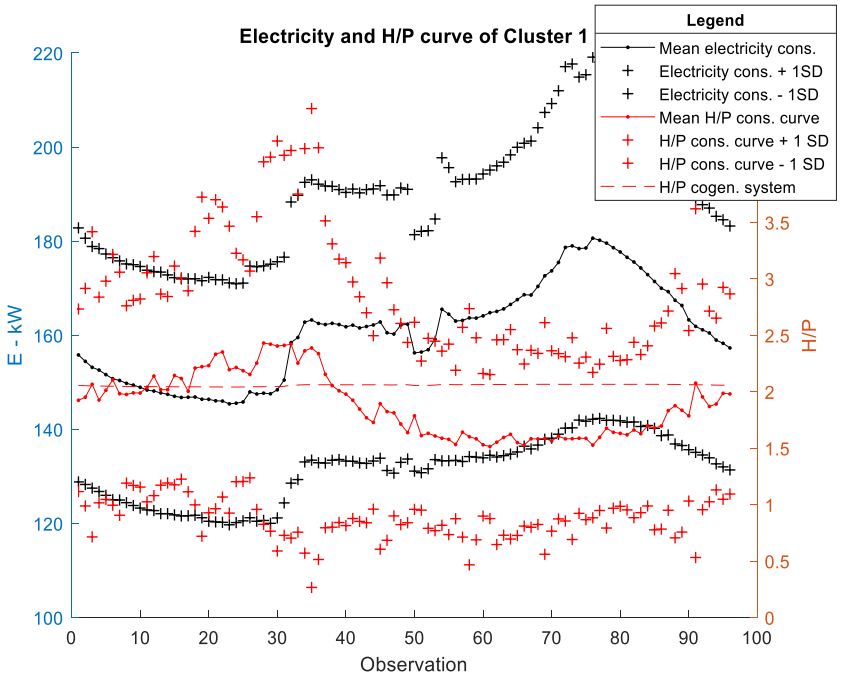


Figure 66. Average electricity power and H/P curves of cluster 1

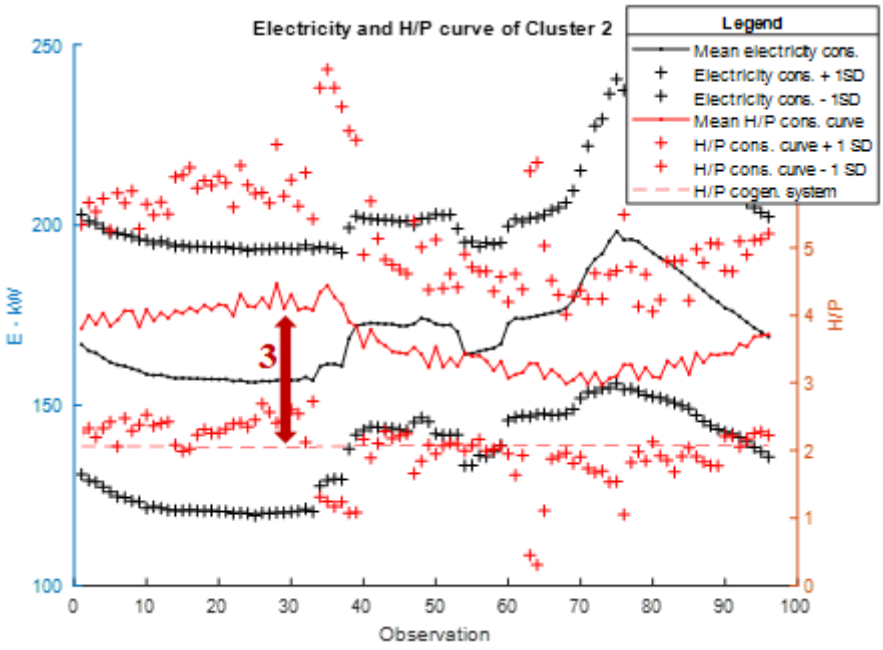


Figure 67. Average electricity power and H/P curves of cluster 2

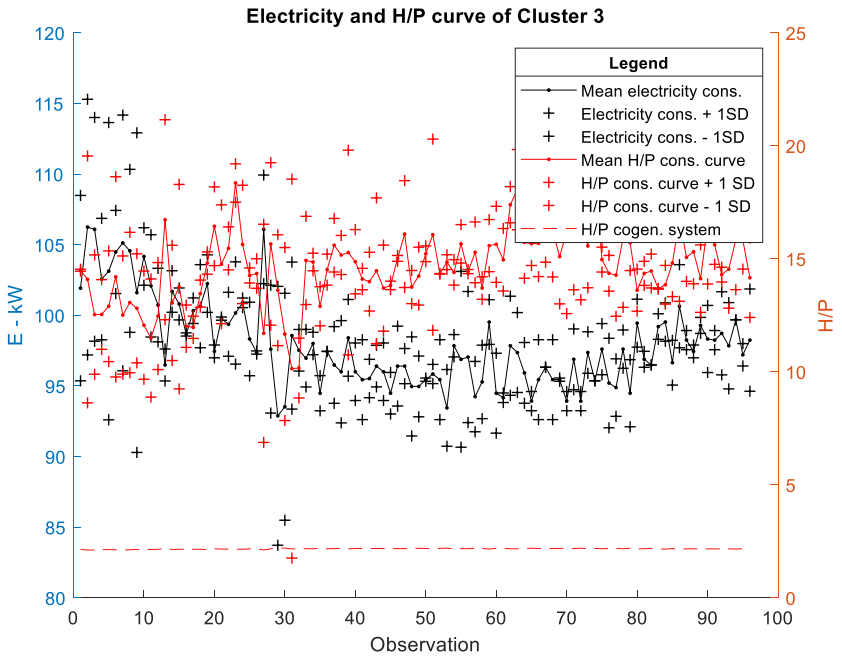


Figure 68. Average electricity power and H/P curves of cluster 3

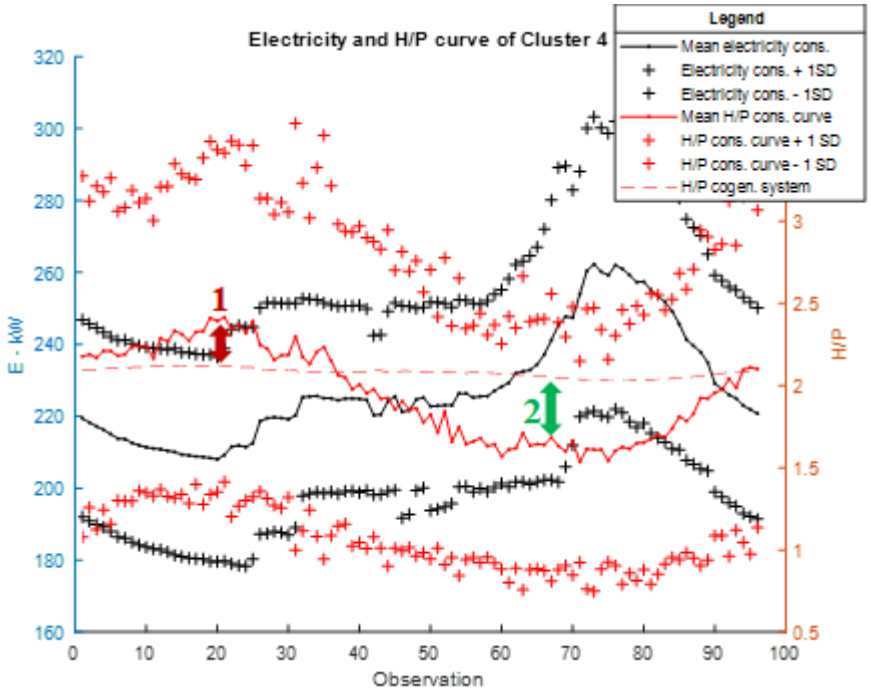


Figure 69. Average electricity power and H/P curves of cluster 4

Figure 69 represents the average electricity power and H/P consumption curves of cluster 4. In this case, the cogeneration system produces less heat than required during the first hours of the day (point 1). Moreover, during the day the mismatch changes, with H/P of the cogeneration system higher than that of the user (point 2). A heat storage is necessary to store the unused heat and to avoid an integration system when demand is higher than generation. The plot suggests also to check the

operation strategy: if the cogeneration system follows only the electricity request, unnecessary heat is available and, even if a storage system is available, total heat produced in an average day is more than the demand. Reference curve of cluster 1 (Figure 66) seems to be similar to reference curve of cluster 4 (Figure 69): both electricity demands and H/P ratio have a similar variation during daytime. It is possible to appreciate that electricity consumption for cluster 1 varies between 130 and 160 kW, meanwhile for cluster 4 it varies between 200 and 250 kW.

Figure 67 represents the average curves for cluster 2. In this case, the cogeneration system is not able to cover the heat demand, and an integration system is necessary (point 3). Moreover, a heat storage system is unnecessary because there is not unused heat during the day.

Cluster 3 (Figure 68) has different reference curves compared to the other clusters: electricity demand and H/P ratio seem to vary differently from cluster 1, cluster 2 and cluster 4. H/P ratio is also higher (between 10 and 15) than the other cases, (maximum is 5). As a matter of fact, this case is related to some observations of Power Analysis (cluster 1, Figure 61 and Figure 62) with high H/P ratio: dataset has just some days (just 0.27% according to Table 20) with higher heat demand than other clusters, and the overall energy request (heat and electricity) is more irregular.

2.4.5 POWER AND PROFILE ANALYSES RESULTS

Results of the Power analysis and the Profile analysis suggest an alternative energy generation system compared to the actual one. As previously mentioned, two ICEs are used as cogenerator, with a heat integration system based on natural gas fired boilers. Water tanks as heat storage are not used. According to the analysis, the most suitable cogeneration technology is a gas micro-turbine with 2 generators of 200 kW nominal electrical power each: each gas turbine has a range of electricity production between 15 kW and 200 kW, meanwhile the H/P ratio can vary between 2.5 (at 30% of partial load) and 2 (at full load).

A model based on polynomial regression is defined according to technical datasheet ([98]). Regression models are created to define electric efficiency and H/P ratio varying the partial load both in electricity driven mode and heat driven mode. Figure 70 represents the gas turbine model where electricity efficiency and H/P ratio are related to the ratio between the net power required and the maximum power: gas turbine is used following electricity demand and partial load is defined on the maximum electric power. Figure 71 represents the gas turbine model where electricity efficiency and H/P ratio are related to the ratio between the heat required and the maximum heat power: gas turbine is used following heat demand, and partial load is defined on the maximum heating power. For all the regressions defined, R^2 is higher than 0.99.

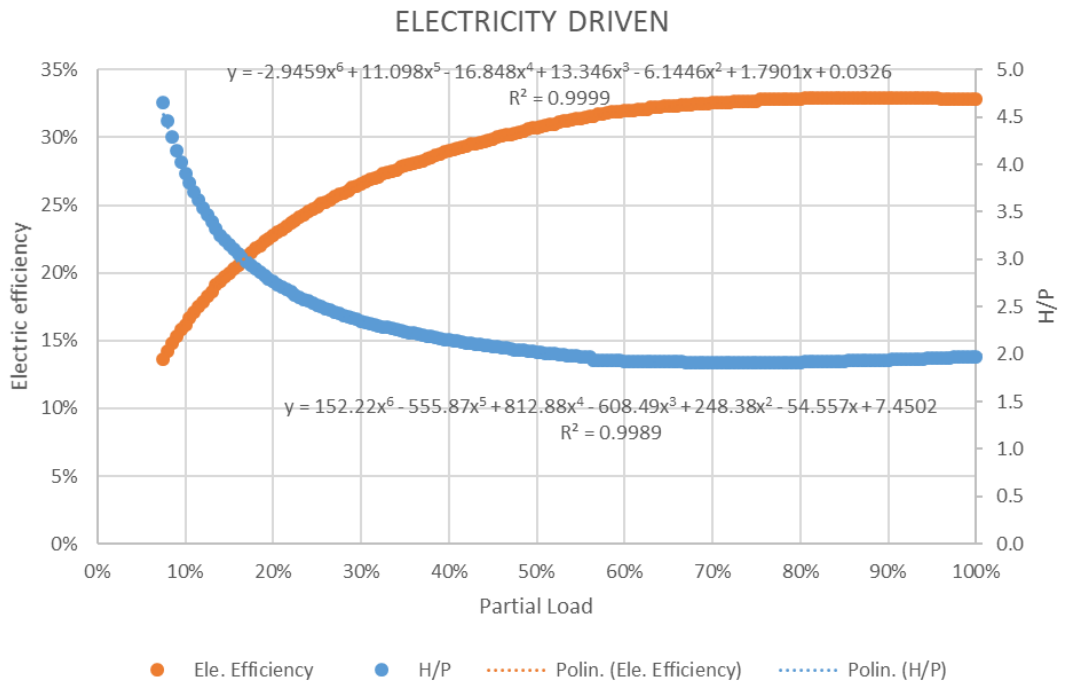


Figure 70. Polynomial regression of electricity efficiency and H/P ratio varying partial load on electricity driven mode

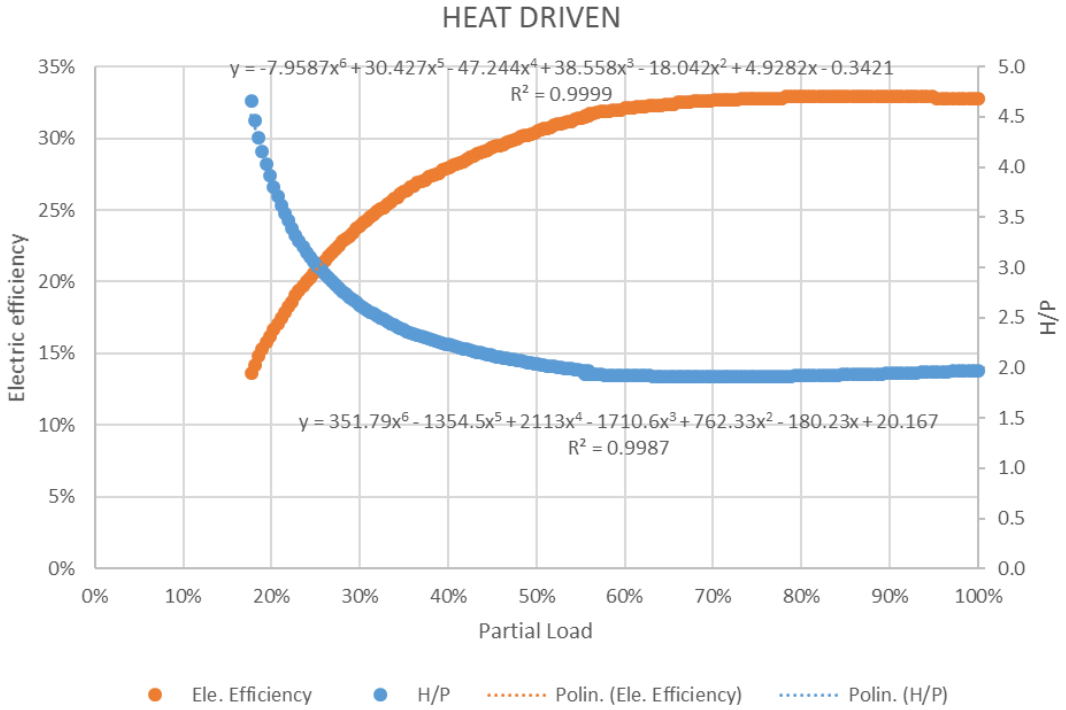


Figure 71. Polynomial regression of electricity efficiency and H/P ratio varying partial load on heat driven mode

Mismatching between the cogeneration system and the energy request occurs, therefore a heat storage system is necessary. A heat storage of 720 kWh is proposed to heat water from the kiln from 20 °C to 75 °C. Such a value is proposed considering the maximum mismatching between the average curves of heat production of the proposed cogeneration system and heat demand. An integration system is required when heat demand is higher than production: Figure 62 shows that in some cases (Power Analysis - Cluster 1) H/P required is higher than H/P available from the cogeneration system proposed. The profile analysis proposed that in some cases (for example Profile analysis – Cluster 2, Figure 67) heat is not available, therefore a heat integration system is required. A natural gas fired boiler with an efficiency of 0.9 is considered as integration system.

Profile analysis suggests not only the opportunity to consider a heat integration and/or a heat storage system, but also the operation strategy. Meanwhile reference curve of cluster 3 (Figure 68) has a higher H/P ratio on consumption than H/P on generation (so an integration is necessary), it happens that for cluster 1 (Figure 66) and cluster 4 (Figure 69) more heat is available than storage, so it is necessary to dissipate it. Consequently, two different scenarios are considered for the operation strategy:

1. Scenario TO BE 1. The cogeneration system follows only the electricity demand;
2. Scenario TO BE 2. The cogeneration system follows the heat demand when both the conditions are satisfied: heat storage is full at 95%, and heat demand is lower than the heat produced in case of operation strategy 1. In the opposite case, the cogeneration system follows the electricity request.

The aim of scenario TO BE 2 is to increase as much as possible the energy efficiency of the system avoiding heat dissipation. On the other hand, electricity consumption from grid increases.

2.4.6 PERFORMANCE OF THE ALTERNATIVE SYSTEM PROPOSED

In this section, an analysis on the performance of the alternative system proposed is performed. Firstly, energy fluxes of the two scenarios TO BE (scenario TO BE 1 and scenario TO BE 2) are proposed in order to analyse how they vary. In particular, we are interested in heat losses (it is supposed that they would be higher for scenario TO BE 1) and electricity consumption from grid (it is supposed that it would be higher for scenario TO BE 2). Successively, a primary energy analysis is performed. The two scenarios are compared with the traditional system (AS IS) by using the primary energy (PE) consumption and the grid electricity exchange as benchmarks. The main scope is to analyse the primary energy saving obtained by changing the cogeneration system and using this operation strategy. The primary energy factor of natural gas is fixed at 9.95 kWh/Sm^3, whereas the electricity exchanged with the grid is considered to be produced with an efficiency of 0.434. Defined F as the total fuel consumption (cogenerator and integration system), $E_{grid,in}$ the total electricity withdrawn from the grid and $E_{grid,out}$ the total electricity exported to the grid, the primary energy function is defined by Eq. 53:

$$PE = F + \frac{E_{grid,in} - E_{grid,out}}{0.434} \qquad \text{Eq. 53}$$

Figure 72 and Figure 73 represent the energy fluxes of the two scenarios. As previously mentioned, if the operation strategy follows only the electricity consumption (TO BE 1 - Figure 72) heat losses occur (they are about 11% of the total heat consumption). The system, however, is grid independent. The hybrid operation strategy of scenario TO BE 2 decreases considerably heat losses, from 0.484 GWh to 0.003 GWh. Though, the consequence is that the electricity withdrawn from the grid increases to 0.222 GWh.

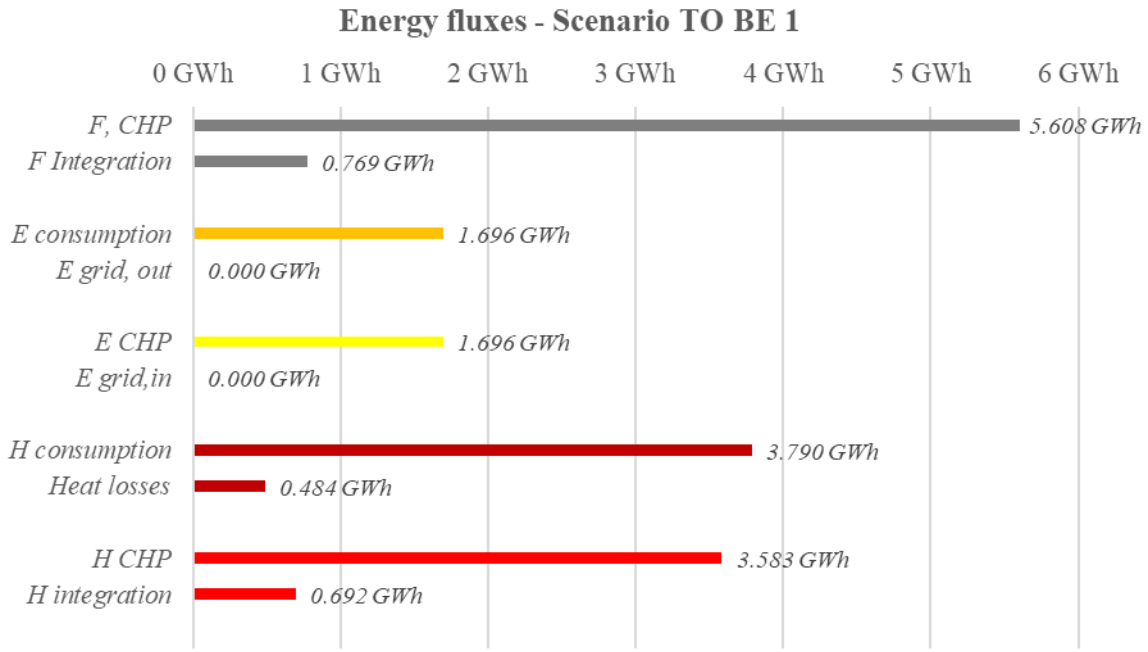


Figure 72. Energy fluxes for Scenario TO BE 1

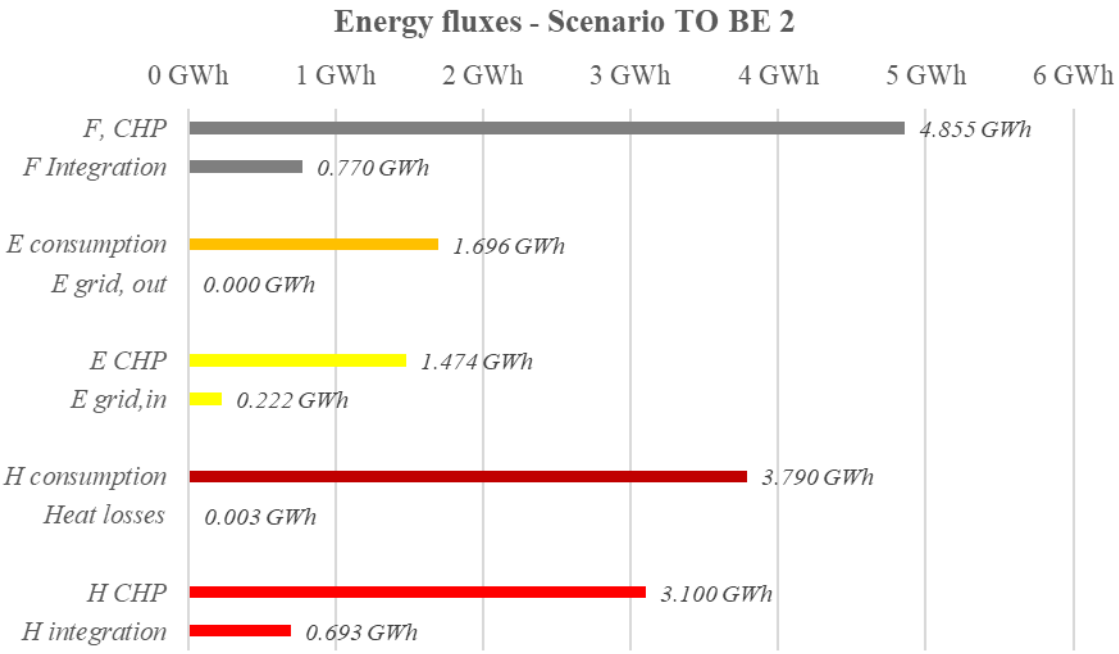


Figure 73. Energy fluxes for Scenario TO BE 2

According to the simulations in both scenarios, saving on primary energy occurs (Table 21). As previously mentioned, the actual cogeneration system (scenario AS IS) uses ICE, that is one of the cogeneration technologies proposed in Power analysis. Gas micro-turbines are chosen for an alternative scenario because the analyses show that a higher H/P ratio is required. The use of gas micro-turbines only (TO BE 1) produces a primary energy saving of 2 % thanks to a better matching between H/P ratio of the user and of the cogenerator. The greatest saving (6 %) is obtained with the hybrid operation system (TO BE 2). The higher energy saving respect to scenario TO BE 1 is a consequence of the operation strategy that significantly decreases heat losses (0.434 GWh for TO BE 1, 0.003 GWh for TO BE 2) even if electricity is consumed from the grid.

Table 21. Primary energy saving of the different scenarios

Scenario	Primary energy	Saving
AS IS	6.505 GWh	-
TO BE 1	6.377 GWh	2.01 %
TO BE 2	6.137 GWh	6.00 %

Heat storage has an important role on the overall efficiency of the system: it is suggested as a consequence of the mismatch proved by the Profile analysis. Without a heat storage, energy demand related to the integration system increases as unused heat of cogeneration system is not stored, so an integration system is required to cover the request. Three benchmarks are proposed to evaluate its influence: index of saving, index of coverage and mean heat stored.

The index of saving (I_S) (Eq. 54) is a benchmark of how much a heat storage prevents that the unused heat is dissipated thus increasing heat losses. It is defined as the ratio between the heat absorbed and stored by the heat storage ($H_{stored,\ in}$) on the total heat produced by cogeneration system (H_{CHP}):

$$I_S = \frac{H_{stored,in}}{H_{CHP}} \qquad \text{Eq. 54}$$

The index of coverage (I_C) (Eq. 55) is proposed as the ratio between total heat stored and used ($H_{stored,\ out}$), and the total heat demand by the user (H_{user}). This index measures the capacity of the heat stored (produced by the cogeneration system when demand is lower than production) to decrease the power of the integration system. It is defined as:

$$I_C = \frac{H_{stored,out}}{H_{user}} \qquad \text{Eq. 55}$$

The mean heat stored in the period is defined as the percentage of the maximum heat storage. Table 22 resumes the results: hybrid operation strategy (TO BE 2) decreases the heat production of the cogenerator system with respect to the electric following strategy (Figure 72 and Figure 73), nevertheless both the percentages of heat stored and the impact of the heat stored to cover heat demand are increased. In particular, the heat stored covers about 5 % on the heat demand in scenario TO BE 2, meanwhile the integration system covers 18 % (0.693 GWh on 3.97 GWh of heat demand). Without heat storage, consumption of the integration system would increase of one third. It is possible to appreciate also that the mean heat storage is close to 50 % (respectively 50.5 % for TO BE 1 and 48.9 % for TO BE 2). Furthermore, even if the mean heat storage is lower in TO BE 2 than TO BE 1, the operation strategy here proposed uses stored heat more efficiently: I_C is 4.68 % compared to 4.31 % of the scenario TO BE 1, meanwhile I_S is 5.73 % and 4.31 % respectively. This means that the heat storage in scenario TO BE 2 has a higher importance both in covering heat demand of the user (I_C benchmark) and in suitably using the heat produced by the cogenerator (I_S benchmark).

Table 22. Benchmark on the heat storage system

Scenario	I_S	I_C	% Mean heat stored
TO BE 1	4.6 %	4.3 %	50.5 %
TO BE 2	5.7 %	4.7 %	48.9 %

2.5 USE OF CLUSTERING AND KNN FOR SHORT FORECASTING

2.5.1 INTRODUCTION

In this chapter a forecast method based on clustering and kNN is proposed. The main goal is to define a methodology to forecast energy request of industrial facilities. Industries use energy (electricity, heating and cooling) both for industrial processes and auxiliaries (lighting, compressed air, etc.). It can be supposed that the energy used for production processes depends on the variety and entity of the production output: if the production output remains constant in terms of type and quantity of items, the energy does not vary significantly.

Machine learning and big data methods have already been proposed to forecast the values of the energy demand from the users and the energy produced by the generation plant. In [99], artificial neural network (ANN) is used to predict residential building energy demand; in [100], support vector machine (SVM) and ANN are applied to predict heat and cooling demand in non-residential sector. In [101], ANN and clustering are used to predict photovoltaic power, whereas in [102] and [103] principal component analysis (PCA) is considered to analyze and forecast photovoltaic data. In [104] and [105], SVM is used.

The aim of this study is to define a model based on machine learning technique that allows to forecast energy demands for a short period (for example the following hour). The method is based on data collected on energy demands, by using a clustering approach. It is supposed that average profiles can be defined using a dataset of observations related to at least one year. Such observations are used to perform the forecast. A method for short forecasting is here proposed. When the energy demand is sampled frequently (for example every 15 minutes) and a dataset is available, data can be used to train a model in order to predict the energy requested. Clustering is used to define average curves, meanwhile kNN (k-nearest neighbors' algorithm) classifies each observation and forecast the energy demand. The clustering method has been already used to classify daily load curves ([106], [107]) and to forecast energy demands ([108], [109]). As a novelty of this study, an innovation on the data normalization to improve performances of the method is proposed.

The first concept introduced is the energy demand curve. It represents a temporal sequence of observations and forecasts of energy demand. Each curve can be split into two parts: support and forecast. The former is the part of the data that will be provided to the model, constituted by the latest observations. Forecast is the predicted data based on the support (Table 23). The length of the support (s) and of the forecast (f) is fixed by the user. In this model, it is proposed that $0 < f \leq s-2$. Section 2.6.5 will illustrate the performance of the model varying f and s for a real case study.

Table 23. Example of curves, definition of support and forecast (sample dataset)

SUPPORT								**FORECAST**			
i = 1	i = 2	…					i = s = 8	j = 1	…		j = f = 4
10	11	10	13	12	14	16	12	11	12	18	13

To perform the forecast, the model features a workflow (Figure 74) based on the following steps:

- **Model training**, a dataset of observations is used to train the model. Observations define average demand curves and train the classification model;
- **Classification**, observations are used to classify which is the most similar average curve;
- **Forecast**, the average curve forecast is used to define forecast of the observations.

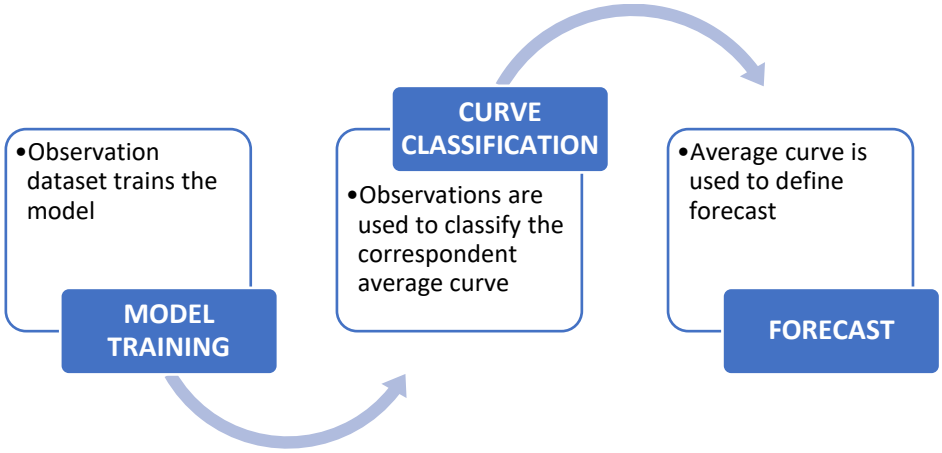


Figure 74. Workflow of the forecast method proposed

The model proposed is based on two machine learning methods: clustering and kNN. Clustering is used only in the training process to define the average curve, whereas kNN is used to classify observations and to relate them with average curves.

2.5.2 MODEL TRAINING

The main task to define the forecast model is the training process. It requires observations of one year or more to train the model properly. Observations are ordered temporarily and then used to defined curves with support and forecast. These curves define a dataset. Workflow of the training (Figure 75) can be divided in the following steps:

1. **Define dataset** – Firstly, it is necessary to define and to normalise a dataset. Then, data are randomly divided into three subgroups: validation, training, and test dataset. These subgroups represent respectively 25 %, 50 % and 25 % of the total observations. The validation dataset is used to define hyperparameters of the model, the training dataset to train both cluster and kNN model and the test dataset to verify the performance of the trained model;
2. **Define hyperparameters** – As previously mentioned, the proposed model defines both cluster and kNN model. Both methods require to define at least the distance function and the number of clusters (cluster model), or the number of observations for classification (kNN model). The euclidean distance function is proposed for cluster model, meanwhile the number of clusters and the number of observations for classification are defined using the validation dataset;

3. **Train cluster model** – When all the hyperparameters are set, the training dataset is used to train the cluster model and to define the average forecast curves;
4. **Train kNN model** – When both the cluster model and consequently the average forecast curves are defined, kNN is defined. kNN is used to forecast observations;
5. **Test model** – The test dataset is used to test the trained model and to check its performances by using MAPE and RMSE.

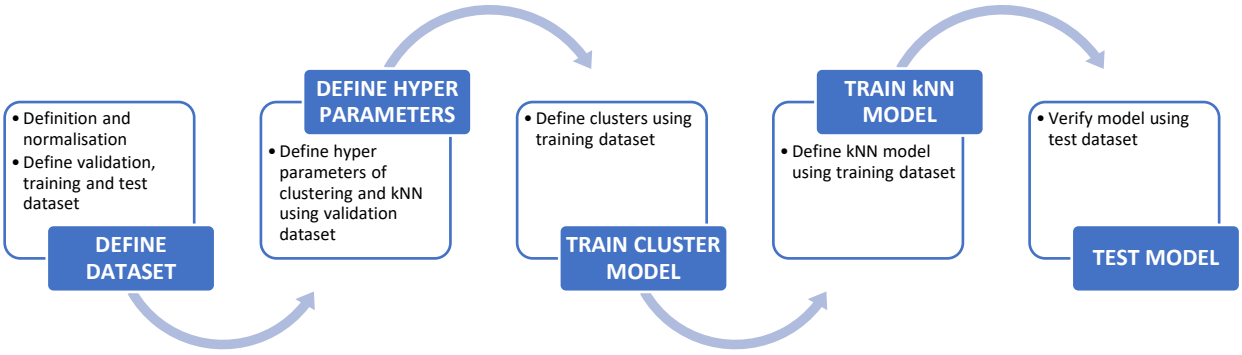


Figure 75. Workflow to train the model

After the training process, the model can be used to forecast new observations.

2.5.3 DATA NORMALIZATION

One of the first step of data analytics is data normalization. As datasets have different values and scale effect may occur, classification methods such as clustering will not work properly if data are not normalized. Usually, normalization is performed using standard score or min-max scaling [69], [110]. Standard score normalizes dataset (X) using the average (μ) and the standard deviation (σ) as described in Eq. 56:

$$\frac{X - \mu}{\sigma} \qquad \text{Eq. 56}$$

In this model, authors propose to normalize dataset differently. As the goal of the model is to forecast energy demand curves, the idea is that different curves may have different scale but similar variation. In this case, the standard score would be normalized but the curves would still have (a lower) scale effect. Instead, in this study it is proposed for each curve to calculate the average of the observations, and then to calculate the variations between observations and average as (Eq. 57):

$$n_{j,i} = \frac{o_{j,i}}{a_j} - 1 \qquad \text{Eq. 57}$$

Where $o_{j,i}$ is the observation i of the curve j, a_j is the average and $n_{j,i}$ is the normalized observation. Figure 76 represents an example explaining why this normalization is proposed. Curves

1 and 2 have different scale but similar variation. Firstly, standard score is applied, then the average normalization. The average (avg) and the standard deviation (std) for the standard score are calculated using all the support values. In the other case, the average of support of each curve is calculated and used for normalization. Forecast values are excluded because they become known only during training process. As it can be seen in Figure 76, curve 2 is 1.58 times bigger than curve 1, and a noise is added. It is possible to appreciate that the proposed method (avg) based on the average of the curves reduces the scale effect, but it keeps the variation. As a matter of fact, normalized curve 1 and 2 have similar values. Instead, the standard score method proposes normalized curves with different values because it normalizes not only the scale effect but the variation as well.

		SUPPORT							FORECAST					
STANDARD SCORE	Curve 1	10.0000	10.1000	10.0000	10.0000	10.1000	10.0000	10.0000	10.0000	10.0000	10.0000	10.0000	avg	12.9369
	Curve 2	15.8000	15.8000	15.9580	15.8000	15.8000	15.8000	15.9580	15.8000	15.8000	15.8000	15.8000	std	3.0187
	norm, curve 1	-0.9729	-0.9398	-0.9729	-0.9729	-0.9398	-0.9729	-0.9729	-0.9729	-0.9729	-0.9729	-0.9729		
	norm, curve 2	0.9485	0.9485	1.0008	0.9485	0.9485	0.9485	1.0008	0.9485	0.9485	0.9485	0.9485		

		SUPPORT							FORECAST					
AVERAGE	Curve 1	10.0000	10.1000	10.0000	10.0000	10.1000	10.0000	10.0000	10.0000	10.0000	10.0000	10.0000	avg, curve 1	10.0286
	Curve 2	15.8000	15.8000	15.9580	15.8000	15.8000	15.8000	15.9580	15.8000	15.8000	15.8000	15.8000	avg, curve 2	15.8451
	norm, curve 1	-0.0028	0.0071	-0.0028	-0.0028	0.0071	-0.0028	-0.0028	-0.0028	-0.0028	-0.0028	-0.0028		
	norm, curve 2	-0.0028	-0.0028	0.0071	-0.0028	-0.0028	-0.0028	0.0071	-0.0028	-0.0028	-0.0028	-0.0028		

Figure 76. Data normalization example

2.5.4 HYPERPARAMETERS DEFINITION

As previously mentioned, it is necessary to define parameters (that are called hyperparameters) for clustering and kNN. Clustering requires the distance function and the number of clusters, kNN requires the number of the nearest neighbors and the distance function. Only the clustering distance function is defined a priori (Euclidean distance), the other ones are defined using validation dataset. Firstly, the number of clusters is defined: as previously mentioned, different criteria have been already developed, they usually try to minimise the number of clusters in order to maximise the distance between data. It is in the author opinion that a more suitable criterion is the minimum number of clusters that minimise the forecasting error. The model here proposed clusters the data to obtain average curves, and then uses them to forecast the energy demand. It is proposed to vary the number of clusters (from 2 to *N*) in a range and for each simulation to calculate MAPE between data and average curves of the clusters. MAPE is the acronym of Mean Absolute Percentage Error and in this case is defined as (Eq. 61):

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}(\frac{1}{l}\sum_{j=1}^{l}(\frac{a_{j,i}}{d_{j,i}} - 1)) \quad \text{Eq. 58}$$

Where n is the number of curves, l is the number of the forecasted values of each curve, $d_{j,i}$ is the value observed and $p_{j,i}$ is the value of the average curve of cluster related to the observed value. The hyperparameter "number of clusters" is then defined as the minimum n that has a MAPE lower than the average of the next three values (Eq. 59):

$$\min(n) \mid MAPE(n) < \frac{MAPE(n+1) + MAPE(n+2) + MAPE(n+3)}{3} \quad \text{Eq. 59}$$

It can be possible to define also n as the minimum number of clusters associated with a MAPE lower than a defined limit (Eq. 60):

$$\min(n) \mid MAPE(n) < MAPE_{limit} \quad \text{Eq. 60}$$

This method can be seen as an early stopping method, because the number of clusters increases as much as the accuracy of the system is increased. Figure 77 and Figure 78 report an example of how this method is applied to a validation dataset of electricity and heat demand, where each curve has 8 observations as support and 4 as forecast. Data refers to the case study defined in section 2.6 and already presented in section 2.4. It is possible to appreciate that curves have a MAPE decreasing rapidly between 2 and 10 clusters, whereas between 10 and 30 they become more stable. With more than 30 clusters the curves have a very low gradient, and locally MAPE increases even if the number of clusters increases. In this case, the method suggests fixing 10 clusters for heat and 13 for electricity.

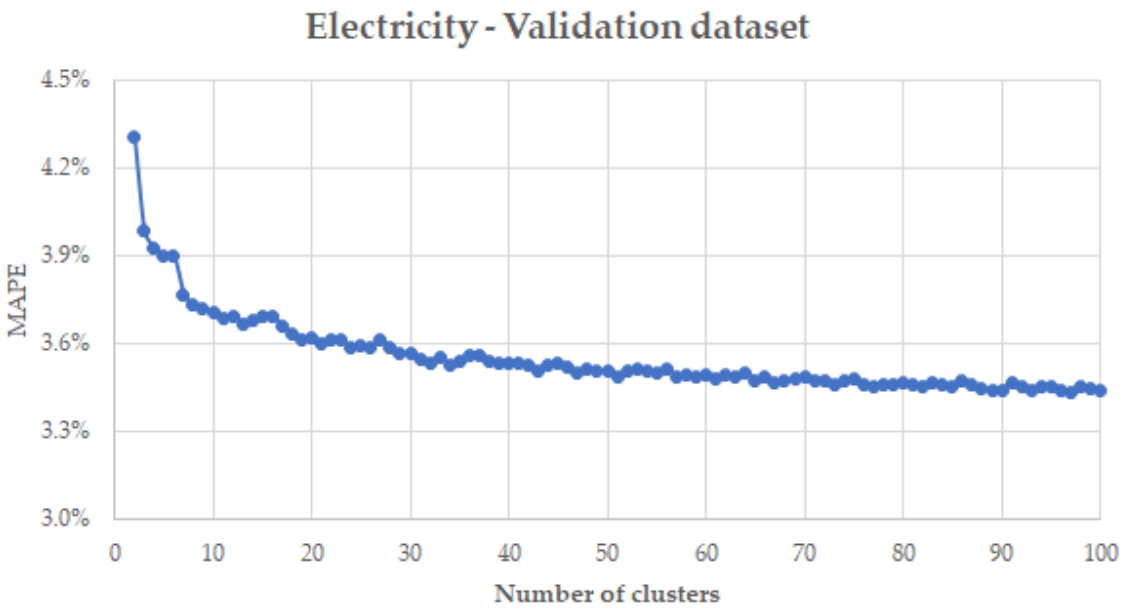


Figure 77. Electricity validation dataset, MAPE varying number of clusters from 2 to 100 – curve 8 - 4

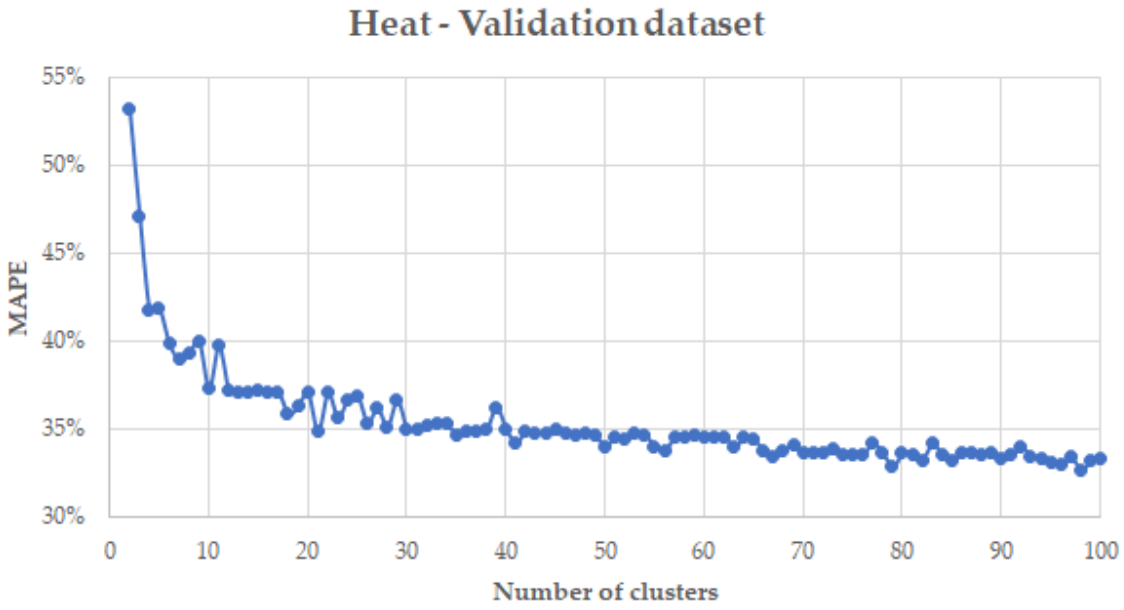


Figure 78. Heat validation dataset, MAPE varying number of clusters from 2 to 100 – curve 8 - 4

As previously mentioned, usually the optimum number of clusters to divide a dataset in is defined using a criterion such as silhouette or gap statistics. Silhouette calculates the average distance between each member of a cluster and another cluster, and the optimum number of clusters is the minimum number that increases distance [92]. If silhouette criterion is applied to the validation dataset (both electricity and heat), the number of clusters that are suggested will be lower than the one that method

proposed: Figure 79 and Figure 80 show that in both cases the suggested number of clusters is two. If this value is used, however, MAPE would be the highest (Figure 77 and Figure 78).

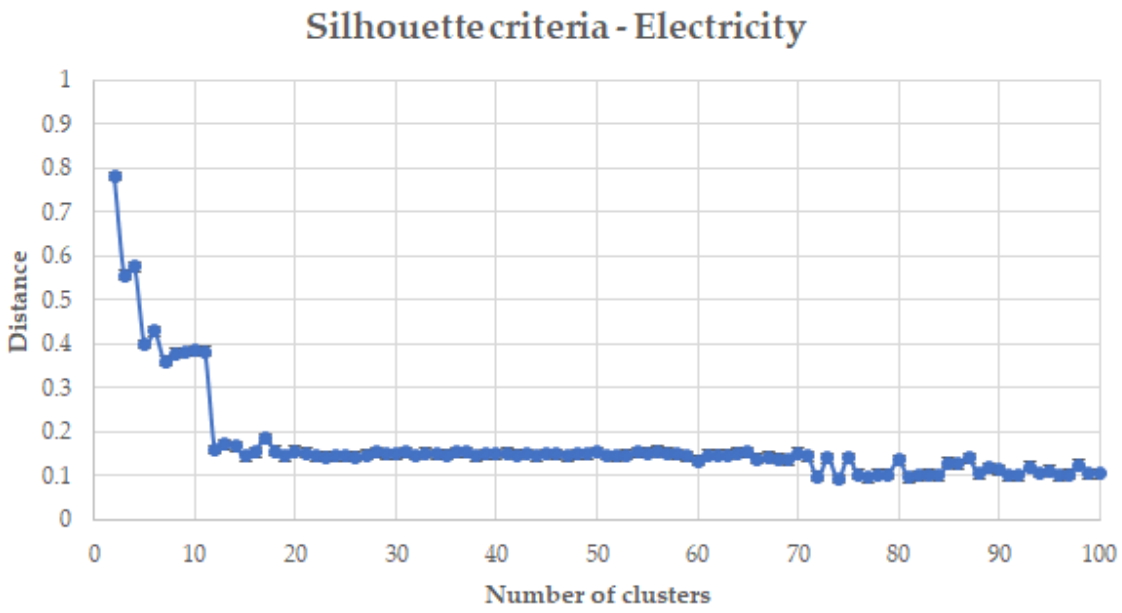


Figure 79. Silhouette applied to electricity validation dataset, distance varying number of clusters from 2 to 100 – curve 8 - 4

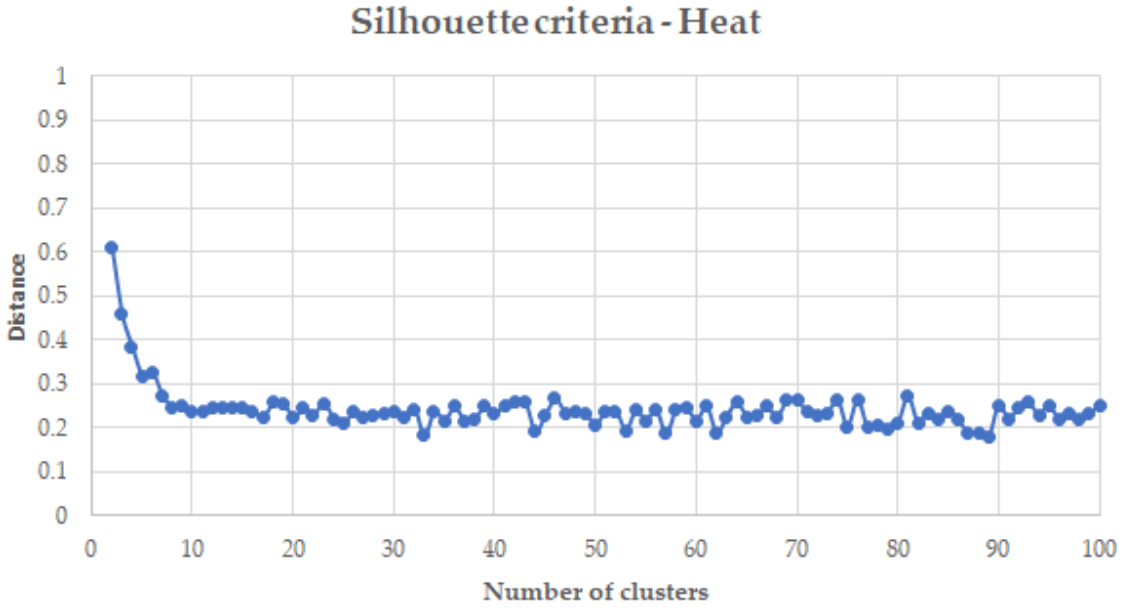


Figure 80. Silhouette applied to heat validation dataset, distance varying number of clusters from 2 to 100 – curve 8 - 4

As previously mentioned, in the proposed model kNN hyperparameters are defined using a MATLAB © optimization function (*Fitchknn*): it optimizes kNN hyperparameters (the distance function and the number of neighbors) to decrease classification error [93].

2.6 SHORT FORECASTING WITH CLUSTERING AND KNN – A CASE STUDY

2.6.1 CASE STUDY INTRODUCTION

The case study presented in section 2.4 is used also in this section to perform forecasting on energy demand. Briefly, the dataset is based on the energy demands of electricity and heat of an industrial facility selling wood (timber) laminated window, plywood, engineered veneer, laminate, flooring and white wood. The industrial process requires heat to dry wood into the kilns (working temperature 70 °C), and to store it into warehouses. Electricity is used for the production equipment, the offices, lighting purpose into the warehouses and to charge electric forklifts. Heat is generated with ICE cogenerators, and a boiler is present as a heat integration system. Figure 81 represents the connection between each component of the system.

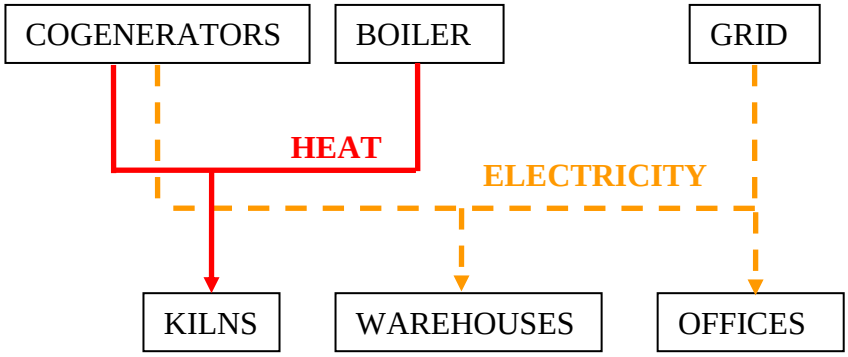


Figure 81. Electricity and heat energy fluxes, connection between generation and utilization

Energy uses (both electricity and heat) were sampled each 15 minutes from 01/01/2015 to 25/09/2017. The electricity demand is available as mean power (kW) in such interval. The heat demand, instead, is calculated by measuring water flow rate (m^3/h) and inlet and outlet temperature (°C) to heat the kilns. Data are stored into a structured SQL database. Data cleaning is performed on the dataset as suggested in section 2.4.2 in order to delete missing measure (NULL value) and/or outliers: an upper and lower limit on electricity and heat request are defined. QQ plot (Figure 82) shows that the dataset with filtered data follows a normal distribution.

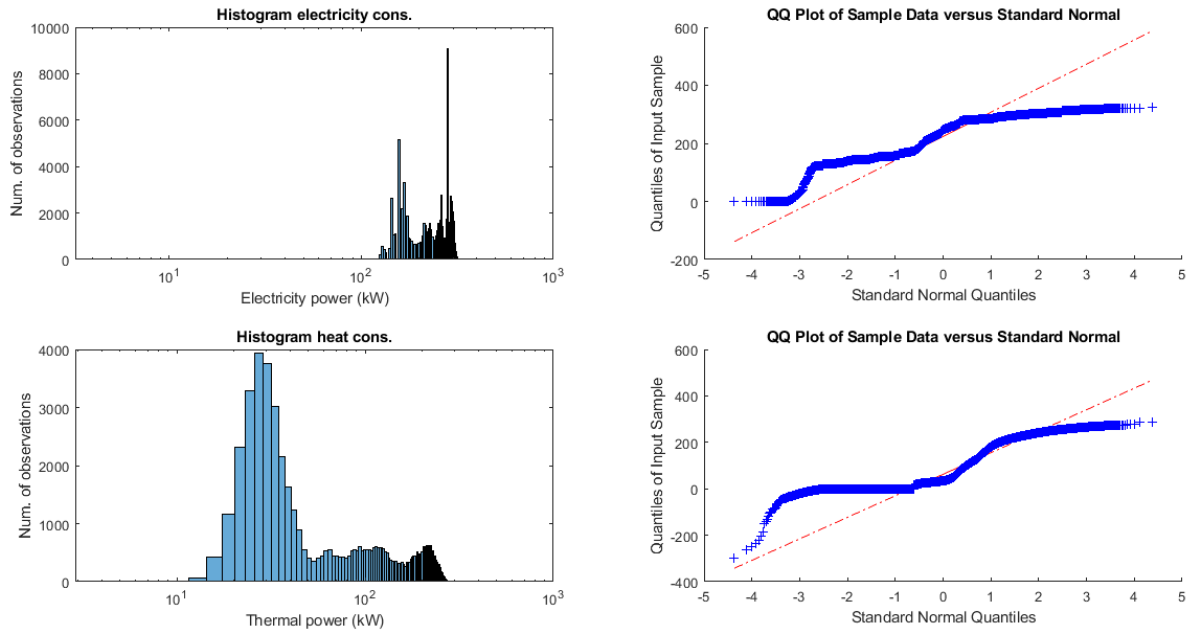


Figure 82. Representation of the dataset filtering data, histogram and QQ plot of electricity demand (top) and heat power demand (bottom)

The proposal is to use these data to define curves with support and forecast to train and validate the forecast model, one for electricity and one for heat. Then an analysis on MAPE and RMSE varying the length of support and forecasting of curves is performed, also it is analyzed the influence of curve size and how normalization impacts on the hyperparameters. Curves with a length of 8-4 (8 values of observation, 4 of forecasting), 10-4, 10-8 and 12-8 are defined for validation and training.

2.6.2 MODEL VALIDATION AND TRAINING

The dataset previously defined is filtered by NULL data or outliers, and split into training, validation and test representing respectively 50 %, 25 % and 25 %. Curves of different lengths for support and forecast are defined in order to discuss the influence of both on the definition of hyperparameters (such the “number of clusters”), in dividing the dataset and on improving the accuracy of the forecast.

Concerning the former, it is necessary to define the number of clusters for each dataset, the number of the nearest neighbors, and the distance function for kNN. The euclidean distance is proposed as the distance function for clustering. Validation is performed as proposed in section 2.5.4. Successively, the model is trained using the training dataset, and then it is tested using the test dataset.

Concerning the accuracy of the forecast, it is defined by calculating MAPE and RMSE between forecasted value and dataset.

2.6.3 ERROR ESTIMATION

When a forecast method is proposed, it is necessary to estimate the error on the forecasting. As previously mentioned, error estimation is used also to define hyperparameters. MAPE and RMSE errors are suggested. Firstly, MAPE is defined as (Eq. 61):

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}(\frac{1}{l}\sum_{j=1}^{l}(\frac{p_{j,i}}{d_{j,i}} - 1)) \quad \text{Eq. 61}$$

Where n is the number of curve, l is the number of the forecasted values of each curve, $p_{j,i}$ is the model predicted value of the curve, and $d_{j,i}$ is the value observed. RMSE is the acronym of Root Mean Square Error. RMSE is proposed instead of the mean square error (MSE) because it is possible to compare the error using the same unit of measurement. It is defined as (Eq. 62):

$$RMSE = \frac{1}{n}\sum_{i=1}^{n}\sqrt{\frac{1}{l}\sum_{j=1}^{l}(p_{j,i} - d_{j,i})^2} \quad \text{Eq. 62}$$

These errors are calculated on the entire forecast, meanwhile the first forecasted value of each curve is the most important. MAPE1 and RMSE1 are calculated considering not all the forecasted values, but only the first one (l = 1).

2.6.4 MODEL TEST

Table 24 resumes some results of the simulations considering energy demand curves of different length as previously mentioned. MAPE is calculated using the test dataset (error between forecasted values and observed values), once for the first forecasted value (Test dataset - MAPE 1) and once for the entire forecast (Test dataset - MAPE). The MAPE value calculated with the validation dataset is reported as well, in order to define the hyperparameter "number of clusters" (section 2.5.4). It is possible to appreciate that "MAPE calculated with validation dataset" is a good predictor of "MAPE calculated with test dataset". For example, considering the "curve 8-4 electricity", the MAPE calculated with the validation dataset is 3.60 %, whereas the MAPE calculated with the test dataset is 3.58 %. Results show also the difference between electricity and heat dataset: curve 8-4 has a MAPE of 3.58 % and 34.11 % respectively. The difference can be explained considering the higher variation of heat values.

Table 24. Simulation of the model with different curves length

Curve	Type of energy	Validation dataset	Test dataset			
		MAPE	MAPE1	MAPE	RMSE1	RMSE
8-4	Electricity	3.60%	2.75%	3.58%	5.15 kW	3.82 kW
8-4	Heat power	35.41%	32.95%	34.11%	93.43 kW	55.43 kW
10-4	Electricity	3.71%	2.74%	3.57%	5.15 kW	3.82 kW
10-4	Heat power	35.23%	32.70%	34.95%	93.20 kW	54.82 kW
10-8	Electricity	4.79%	2.90%	4.47%	5.47 kW	3.53 kW
10-8	Heat power	36.66%	35.30%	34.12%	90.03 kW	41.99 kW
12-8	Electricity	4.69%	2.80%	4.47%	5.31 kW	3.53 kW
12-8	Heat power	39.00%	32.10%	37.21%	95.14 kW	43.05 kW

2.6.5 INFLUENCE OF THE CURVE SIZE

Observations are used to define the curves to train and to test the forecast model. Support is the part of the curve that is used to classify observation and, consequently, it defines the forecasted value (forecast part). The length of support (*s*) and forecast (*f*) may vary the hyperparameter "number of clusters" and, consequently, the error on forecast. Increasing forecast length (with same support length) is expected to increase forecast error because the model needs to predict more observations. It is unknow if the effect of increasing the support length (with the same forecast length) could increase or decrease the accuracy on the curve classification. Figure 83 and Figure 84 represent the value of MAPE criteria for the validation dataset, varying support and forecast for electricity and heat respectively.

Firstly, it is possible to appreciate that the electricity validation dataset has more regular variation of MAPE than the heat validation dataset. With the electricity dataset, MAPE increases by increasing support and/or forecast lengths: it is supposed that the electricity demand varies differently from the heat. As expected, electricity dataset shows that the MAPE increases by increasing forecast length of the curve. The MAPE increases from 3.5 % of 16-2 curve (4 support length, 2 forecast length) to 6.3 % of 16-4 curve: the more forecasts are required, the more the error increases. On the other hand, the increase of support length is related to the increase of the MAPE as well: from 2.9 % of curve 4-2 to 3.5 % of curve 16-2. Even if more observations are available to classify each curve, error does not decrease.

	Electricity	FORECAST						
		2	4	6	8	10	12	14
SUPPORT	4	2.9%						
	6	2.9%	3.6%					
	8	3.1%	3.6%	4.1%				
	10	3.1%	3.7%	4.1%	4.8%			
	12	3.2%	3.8%	4.3%	4.7%	5.1%		
	14	3.3%	3.9%	4.4%	4.9%	5.3%	5.8%	
	16	3.5%	4.0%	4.5%	5.1%	5.3%	5.8%	6.3%

Figure 83. Heatmap of MAPE of electricity validation dataset with curves with different support and forecast length

	Heat Power	FORECAST						
		2	4	6	8	10	12	14
SUPPORT	4	37.7%						
	6	36.7%	35.2%					
	8	33.8%	35.4%	39.2%				
	10	36.8%	35.2%	37.3%	36.7%			
	12	40.2%	37.8%	39.0%	39.0%	36.9%		
	14	35.0%	36.9%	33.1%	36.1%	39.2%	37.4%	
	16	35.4%	39.0%	37.8%	39.0%	39.3%	37.9%	38.8%

Figure 84. Heatmap of MAPE of heat power validation dataset with curves with different support and forecast length

2.6.6 INFLUENCE OF THE NORMALIZATION

As mentioned in section 2.5.3, here we propose a normalization based on the percentage norm instead of the standard score. The aim is to reduce the scale effect of the curves and to maintain their variation. Figure 85 and Figure 86 represent the MAPE varying the number of clusters on the electricity validation dataset. A curve of 8 observations as support and 4 for forecast is used in Figure 85, and one of 10 observations for support and 4 for forecast is used in Figure 86. In both cases, dataset has a higher MAPE when it is normalized with standard score rather than with the percentage norm proposed.

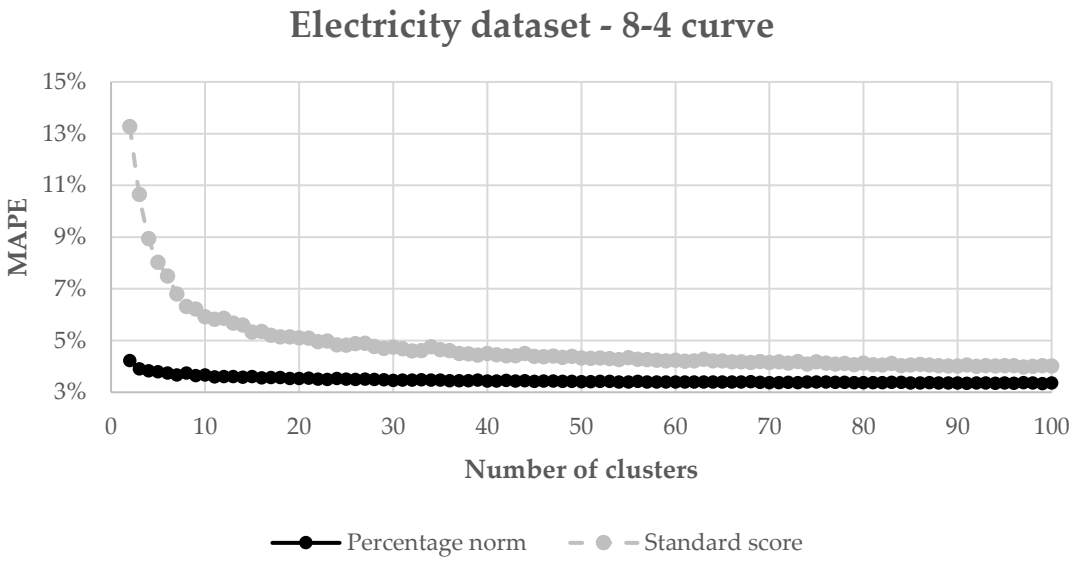


Figure 85. Comparison on MAPE with the electricity validation dataset, curve with 8 observation and 4 forecast values, normalisation between percentage norm and standard score

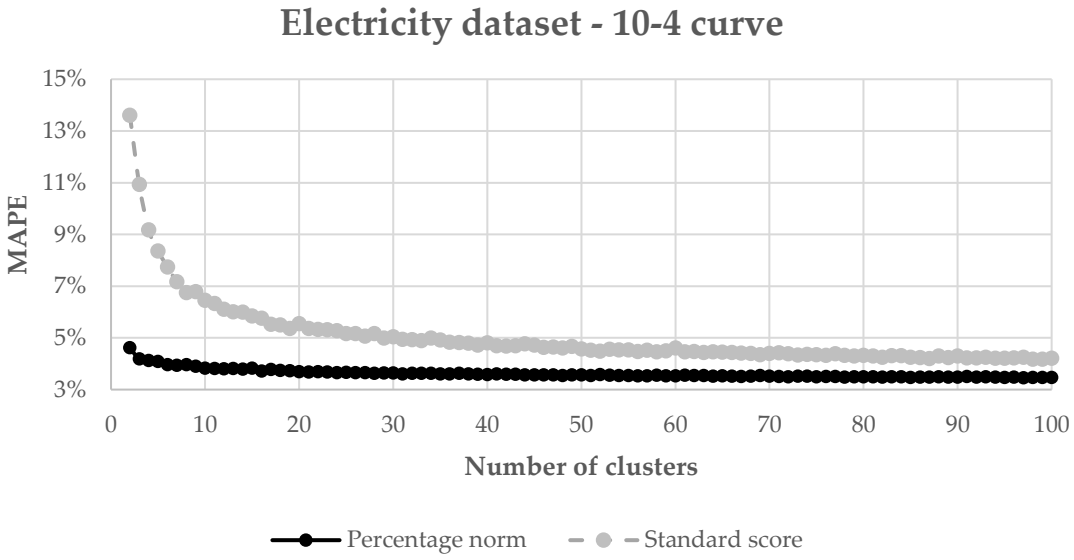


Figure 86. Comparison on MAPE with the electricity validation dataset, curve with 10 observation and 4 forecast values, normalisation between percentage norm and standard score

2.6.7 CONCLUSIONS

In this chapter, a model for short forecasting is proposed and analyzed. As in the authors' opinion this task is a pattern recognition problem: machine learning methods such as "clustering analysis" and "kNN" are proposed to perform the task. A dataset of observations is required to define hyperparameters, to train the model and to test it. Novelties on hyperparameters ("number of clusters" definition) and on data normalization are proposed to increase the performances of the method. A case study is presented in order to apply the proposed method, to analyze how the length of energy demand curves (numbers of observations and forecast) impacts on the model, and to check its

performances. Results show that the improvements here proposed increase the performance, meanwhile the length of the curves (both on support and forecast) affects error: the higher the length (both on support and/or forecast), the higher the error. A validation dataset is not used only to define hyperparameters, but also to predict error on forecast. It is in the authors' opinion that further improvements could be achieved by studying the most suitable distance function for the dataset and/or by weighting the observations.

CONCLUSIONS

Industrial sector is one of the main consumers of energy that is based mainly on hydrocarbons; it directly contributes to greenhouse gases emissions and, consequently, to global warming. Improvements on energy generation systems for industrial facilities and their management are required to decrease the environmental impact. In this thesis, innovative energy generation systems are proposed firstly, based on novel technologies based on SOFC and RSOC. Successively, machine learning methods are applied to perform data analytics and artificial intelligence techniques on energy data in order to define energy system with higher efficiency and to perform forecasting with a high accuracy.

SOFC (and also RSOC) are proposed as energy generator for industrial systems thanks to their advantages: possibility of using different types of fuel (from natural gas to low weight hydrocarbons such as butane or propane), heat available at high temperature, modularity and high efficiency on energy conversion (both electricity and overall efficiency).

An innovative heat recovery system SOFC with air source heat pump is proposed, where SOFC waste gases heat up air at the heat pump evaporator inlet to increase the coefficient of performance. Simulations with different temperatures and relative humidity were performed, and the results show that the heat pump performance increases up to 100% whit high humidity and/or high SOFC utilization. Primary energy saving between 37.5% and 45% are reported comparing the proposed system with separate energy production.

SOFC combined with RSOC are then analyzed as a flexible energy system where it is necessary to vary heat to power ratio: the aim is to match H/P between energy generation and demand in order to avoid heat integration system with a lower efficiency on energy conversation. A case study based on a paper mill is presented to analyze the possibility of revamping the energy generation system by using SOFC/RSOC and dismissing the old steam turbine. The main aim is to match H/P between energy generation and demand with a higher efficiency, and to produce hydrogen as sub-product of the system. Results of the simulation show that is possible to achieve a primary energy saving up to 6%: the higher the SOFC power, the higher the achieved PES. Hydrogen production could reach a production rate of 16 kg/h but exclusively when only one paper production line is in operation. In the whole range of variation of the investment cost (-10% to +30% of the value proposed in literature), the RSOC integrated system has a lower cost for hydrogen generation than the traditional system actually proposed.

Improvements on industrial energy system could be achieve not only with innovative energy components, but also using energy consumption data with artificial intelligence and machine learning. Clustering is suggested as machine learning method for data analytics, to define the most suitable energy generation technology and to forecast energy consumption. Two innovative analyses based on clustering of energy consumption data (power and profile analyses) are illustrated in the thesis. These analyses divide data into homogenous groups (cluster), to define firstly which is the most

suitable polygeneration technology (power analysis) and then to define average days of consumption (profile analysis) to identify the most suitable operation strategy. These methods are then applied to a case study. Power analysis on energy data sampled every 15 minutes for 2 years suggests the most suitable cogeneration system (micro gas turbine) and the nominal power. Then profile analysis is used to check the operation strategy minimizing energy losses and increasing overall efficiency suggesting a heat storage system and its capacity.

Finally, clustering combined with k-nearest neighbors (kNN) is proposed to find similar pattern of energy demand, to identify average demand profiles and then to use them to forecast energy request of the next hours. Even if clustering combined with kNN have been already analyzed as forecasting methods, in this thesis novelties on hyperparameters definition and data normalization are proposed. Firstly, it is suggested to normalize data before clustering using a percentage norm instead of a norm based on min-max or using mean and standard deviation of the dataset: the aim is to decrease only scale effect on the data, not its variation. Then the number of cluster hyperparameter is set using error estimation function (for example mean absolute percentage error) between data observed and average curves instead of silhouette and/or gap statistic criteria as suggested as state-of-art: the proposal is to know a priori which the forecast error would be. The methodology is then applied to the case study previously analyzed and electricity and heat consumptions are then forecasted.

POSTFACE

At time of writing (26/11/2019) parts of the thesis were already published or are under review.

Co-Authors	Journal	Title	DOI
Vialetto Giulio, Noro Marco, Rokni Masoud	Proceedings "12th SDEWES Conference", Dubrovnik, 2017, SDEWES2017.75, ISSN 1847-7178	Analysis of a cogeneration system based on solid oxide fuel cell and air source heat pump with novel heat recovery	Not applicable
Vialetto Giulio, Noro Marco, Rokni Masoud	Journal of Electrochemical Energy Conversion and Storage, 2019, 16(2), 021005, Paper No: JEECS-18-1064	Studying a hybrid system based on solid oxide fuel cell combined with an air source heat pump and with a novel heat recovery	https://doi.org/10.1115/1.4041864
Vialetto Giulio, Noro Marco, Colbertaldo Paolo	International Journal of Hydrogen Energy, 2019, 44(19), pp. 9608-9620	Enhancement of energy generation efficiency in industrial facilities by SOFC – SOEC systems with additional hydrogen production	https://doi.org/10.1016/j.ijhydene.2018.08.145
Vialetto Giulio, Noro Marco	Proceedings "14th SDEWES Conference", Dubrovnik, 2019	An innovative approach to design cogeneration systems based on big data analysis and use of clustering methods	Not applicable
Vialetto Giulio, Noro Marco	Energies, 2019, 12(23), 4407	Short forecasting method based on clustering and kNN: application to an industrial facility powered by a cogenerator	https://doi.org/10.3390/en12234407
Vialetto Giulio, Noro Marco	Energy Conversion & Management	An innovative approach to design cogeneration systems based on big data analysis and use of clustering methods (*under review*)	Not applicable

Table P. 1. List of articles already published or under review regarding the following thesis (Last update 26/11/2019)

Other articles published during Ph.D. and not related to thesis and/or published before starting the Ph.D.

Co-Authors	Journal	Title	DOI
Vialetto Giulio, Noro Marco, Rokni Masoud	International Journal of Hydrogen Energy, 2017 42(15), pp. 10285-10297	Combined micro-cogeneration and electric vehicle system for household application: An energy and economic analysis in a Northern European climate	https://doi.org/10.1016/j.ijhydene.2017.01.035
Vialetto Giulio, Noro Marco, Rokni Masoud	Journal of Sustainable Development of Energy, Water and Environment Systems, 2017, 5(4), pp. 590-607	Thermodynamic Investigation of a Shared Cogeneration System with Electrical Cars for Northern Europe Climate	http://dx.doi.org/10.13044/j.sdewes.d5.0162

Co-Authors	Journal	Title	DOI
Vialetto Giulio, Rokni Masoud	Renewable Energy, 2015, 78, pp. 146-156	Innovative household systems based on solid oxide fuel cells for a northern European climate	https://doi.org/10.1016/j.renene.2015.01.012
Vialetto Giulio, Rokni Masoud	Proceedings of the Global Conference on Global Warming 2015	A new cogeneration residential system based on solid oxide fuel cells a northern European climate	Web reference is : https://orbit.dtu.dk/en/publications/a-new-cogeneration-residential-system-based-on-solid-oxide-fuel-cells-for-a-northern-european-climate(c755ba4e-b290-41fa-a876-b7d231fb2c74).html
Rokni, Masoud ; Vialetto, Giulio	Proceedings of The 10th International Green Energy Conference, IGEC-2015-1367	Residential Systems Based on Solid Oxide Fuel Cells for Scandinavian Climate	Web reference is: https://orbit.dtu.dk/en/publications/residential-systems-based-on-solid-oxide-fuel-cells-for-scandinavian-climate(67aa39da-8b13-4c60-b2e0-b4c704b39658).html
Vialetto Giulio, Noro Marco, Rokni Masoud	International Journal of Hydrogen Energy, 2015, 40(41), pp. 14378-14391	Innovative household systems based on solid oxide fuel cells for the Mediterranean climate	http://dx.doi.org/10.1016/j.ijhydene.2015.03.085
Vialetto Giulio, Noro Marco, Rokni Masoud	Proceedings “2nd SEE SDEWES Conference”, Piran, 2016, SDEWES.SEE2016.0248, ISSN 1847-7178	Thermodynamic analysis of a shared cogeneration system and electrical mobility located in a Northern Europe climate	Not applicable

Table P. 2. List of articles published during Ph.D. and not related to thesis and/or before starting the Ph.D.

REFERENCES

[1] Eurostat, “Energy statistics - An overview,” 2018. [Online]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Energy_statistics_-_an_overview#Final_energy_consumption. [Accessed: 07-Jul-2019].

[2] A. Beyene and A. Moman, “Process oriented industrial classification based on energy intensity,” *Appl. Therm. Eng.*, vol. 26, no. 17–18, pp. 2079–2086, Dec. 2006.

[3] EPA, *Cataloug of CHP Technologies*. EPA - Enviroment Protection Agency, 2017.

[4] EG&G Technical Services Inc.; U.S. Department of Energy (DOE), “Fuel Cell Handbook (7th Edition),” 2004.

[5] P. Jasinski, “Micro solid oxide fuel cells and their fabrication methods,” *Microelectron. Int.*, vol. 25, no. 2, pp. 42–48, Apr. 2008.

[6] B. Elmegaard, *Simulation of Boiler Dynamics: Development, Evaluation and Application of a General Energy System Simulation Tool*. 1999.

[7] M. Rokni, “Plant characteristics of an integrated solid oxide fuel cell cycle and a steam cycle,” *Energy*, vol. 35, no. 12, pp. 4691–4699, Dec. 2010.

[8] M. Rokni, “Thermodynamic analysis of SOFC (solid oxide fuel cell)–Stirling hybrid plants using alternative fuels,” *Energy*, vol. 61, pp. 87–97, Nov. 2013.

[9] M. Rokni, “Addressing fuel recycling in solid oxide fuel cell systems fed by alternative fuels,” *Energy*, vol. 137, pp. 1013–1025, Oct. 2017.

[10] G. Vialetto, M. Noro, and M. Rokni, “Thermodynamic investigation of a shared cogeneration system with electrical cars for northern Europe climate,” *J. Sustain. Dev. Energy, Water Environ. Syst.*, vol. 5, no. 4, 2017.

[11] A. Mazzucco and M. Rokni, “Thermo-economic analysis of a solid oxide fuel cell and steam injected gas turbine plant integrated with woodchips gasification,” *Energy*, vol. 76, pp. 114–129, Nov. 2014.

[12] M. Rokni, “Biomass gasification integrated with a solid oxide fuel cell and Stirling engine,” *Energy*, vol. 77, pp. 6–18, Dec. 2014.

[13] M. Rokni, “Thermodynamic analyses of municipal solid waste gasification plant integrated with solid oxide fuel cell and Stirling hybrid system,” *Int. J. Hydrogen Energy*, vol. 40, no. 24, pp. 7855–7869, Jun. 2015.

[14] N. C. Ullvius and M. Rokni, “A study on a polygeneration plant based on solar power and solid oxide cells,” *Int. J. Hydrogen Energy*, vol. 44, no. 35, pp. 19206–19223, Jul. 2019.

[15] G. Vialetto and M. Rokni, “Innovative household systems based on solid oxide fuel cells for a northern European climate,” *Renew. Energy*, vol. 78, 2015.

[16] G. Vialetto, M. Noro, and M. Rokni, “Innovative household systems based on solid oxide fuel cells for the Mediterranean climate,” *Int. J. Hydrogen Energy*, vol. 40, no. 41, pp. 14378–14391, Nov. 2015.

[17] G. Vialetto, M. Noro, and M. Rokni, “Combined micro-cogeneration and electric vehicle system for household application: An energy and economic analysis in a Northern European climate,” *Int. J. Hydrogen Energy*, vol. 42, no. 15, 2017.

[18] G. Vialetto, M. Noro, and M. Rokni, “Studying a Hybrid System Based on Solid Oxide Fuel Cell Combined With an Air Source Heat Pump and With a Novel Heat Recovery,” *J. Electrochem. Energy Convers. Storage*, vol. 16, no. 2, p. 021005, Dec. 2018.

[19] G. Vialetto, M. Noro, P. Colbertaldo, and M. Rokni, “Enhancement of energy generation efficiency in industrial facilities by SOFC – SOEC systems with additional hydrogen production,” *Int. J. Hydrogen Energy*, Sep. 2018.

[20] IETS, “Application of Industrial Heat Pumps.” [Online]. Available: https://iea-industry.org/app/uploads/annex-xiii-part-a.pdf.

[21] A. Kapil, I. Bulatov, R. Smith, and J.-K. Kim, “Process integration of low grade heat in process industry with district heating networks,” *Energy*, vol. 44, no. 1, pp. 11–19, Aug. 2012.
[22] H. Fang, J. Xia, and Y. Jiang, “Key issues and solutions in a district heating system using low-grade industrial waste heat,” *Energy*, vol. 86, pp. 589–602, Jun. 2015.
[23] H. Fang, J. Xia, K. Zhu, Y. Su, and Y. Jiang, “Industrial waste heat utilization for low temperature district heating,” *Energy Policy*, vol. 62, pp. 236–246, Nov. 2013.
[24] A. N. Ajah, A. C. Patil, P. M. Herder, and J. Grievink, “Integrated conceptual design of a robust and reliable waste-heat district heating system,” *Appl. Therm. Eng.*, vol. 27, no. 7, pp. 1158–1164, May 2007.
[25] K. Holmgren, “Role of a district-heating network as a user of waste-heat supply from various sources – the case of Göteborg,” *Appl. Energy*, vol. 83, no. 12, pp. 1351–1367, Dec. 2006.
[26] UNI, “UNI/TS 11300-4:2010 ‘Renewable energy and other generation systems for space heating and domestic hot water production.’” 2010.
[27] EN, “EN 14825:2008 ‘Air conditioners, liquid chilling packages and heat pumps, with electrically driven compressors, for space heating and cooling - Testing and rating at part load conditions and calculation of seasonal performance.’” 2008.
[28] F. Busato, R. M. Lazzarin, and M. Noro, “Energy and economic analysis of different heat pump systems for space heating,” *Int. J. Low-Carbon Technol.*, vol. 7, no. 2, pp. 104–112, Jun. 2012.
[29] Viessmann, “Techinical datasheet Vitocal 200-A.” [Online]. Available: http://downloads.viessmannitalia.it/documents/5820-437-052015.pdf. [Accessed: 12-Feb-2016].
[30] Busato F., Lazzari, R., Minchio F., Noro M., “Frosting factor,” in *Sorgenti termiche delle pompe di calore. Aspetti tecnici, economici e normativi (Heat sources of heat pumps. Technical, economic and standard aspects, in Italian)*, 2012.
[31] ASHRAE, “Chapter 1 – Psychometrics,” in *ASHRAE Handbook—Fundamentals (SI)*, 2009.
[32] European Parlament, “Directive 2012/27/EU - Energy efficiency.” 2012.
[33] European Parlament, “Directive 2004/8/EC - Promotion of cogeneration based on a useful heat demand in the internal energy market.” 2004.
[34] C. Acar and I. Dincer, “Comparative assessment of hydrogen production methods from renewable and non-renewable sources,” *Int. J. Hydrogen Energy*, vol. 39, no. 1, pp. 1–12, Jan. 2014.
[35] D. Astiaso Garcia, F. Barbanera, F. Cumo, U. Di Matteo, and B. Nastasi, “Expert Opinion Analysis on Renewable Hydrogen Storage Systems Potential in Europe,” *Energies*, vol. 9, no. 11, p. 963, Nov. 2016.
[36] G. Guandalini, S. Campanari, and M. C. Romano, “Power-to-gas plants and gas turbines for improved wind energy dispatchability: Energy and economic assessment,” *Appl. Energy*, vol. 147, pp. 117–130, Jun. 2015.
[37] B. Castellani, S. Rinaldi, E. Morini, B. Nastasi, and F. Rossi, “Flue gas treatment by power-to-gas integration for methane and ammonia synthesis – Energy and environmental analysis,” *Energy Convers. Manag.*, vol. 171, pp. 626–634, Sep. 2018.
[38] E. S. with E. E. L. for the F. C. and H. J. Undertaking, “Fuel cells and hydrogen Joint undertaking, Development of Water Electrolysis in the European Union.” [Online]. Available: http://www.fch.europa.eu/. [Accessed: 06-Feb-2018].
[39] A. Buttler and H. Spliethoff, “Current status of water electrolysis for energy storage, grid balancing and sector coupling via power-to-gas and power-to-liquids: A review,” *Renew. Sustain. Energy Rev.*, vol. 82, pp. 2440–2454, Feb. 2018.
[40] M. Rokni, “Analysis of a polygeneration plant based on solar energy, dual mode solid oxide cells and desalination,” *Int. J. Hydrogen Energy*, vol. 44, no. 35, pp. 19224–19243, Jul. 2019.
[41] M. NI, M. LEUNG, and D. LEUNG, “Energy and exergy analysis of hydrogen production by

solid oxide steam electrolyzer plant," *Int. J. Hydrogen Energy*, vol. 32, no. 18, pp. 4648–4660, Dec. 2007.
[42] M. Ni, M. K. H. Leung, and D. Y. C. Leung, "A modeling study on concentration overpotentials of a reversible solid oxide fuel cell," *J. Power Sources*, vol. 163, no. 1, pp. 460–466, Dec. 2006.
[43] E. Hernández-Pacheco, D. Singh, P. N. Hutton, N. Patel, and M. D. Mann, "A macro-level model for determining the performance characteristics of solid oxide fuel cells," *J. Power Sources*, vol. 138, no. 1–2, pp. 174–186, Nov. 2004.
[44] A. G. Power, "Alto Garda Power - Cogeneration for paper mill combined with district heating." [Online]. Available: http://www.altogardapower.eu/. [Accessed: 21-Jul-2019].
[45] U. S. D. of Energy, "Gas Turbine." [Online]. Available: http://www.publicdomainfiles.com/show_file.php?id=14019791416948. [Accessed: 28-Jul-2019].
[46] Margen, "ICE." [Online]. Available: http://www.margen.it/en/catalog/mps-cogeneration. [Accessed: 28-Jul-2019].
[47] Holland - RET, "Micro Gas Turbine." [Online]. Available: https://holland-ret.com/flex-turbine-mt250-250-kw/. [Accessed: 28-Jul-2019].
[48] E. T. Center, "Steam Turbines." [Online]. Available: https://edisontechcenter.org/steamturbines.html. [Accessed: 28-Jul-2019].
[49] N. Aden, "Necessary but not sufficient: the role of energy efficiency in industrial sector low-carbon transformation," *Energy Effic.*, vol. 11, no. 5, pp. 1083–1101, Jun. 2018.
[50] D. Chiaroni, M. Chiesa, V. Chiesa, S. Franzò, F. Frattini, and G. Toletti, "Introducing a new perspective for the economic evaluation of industrial energy efficiency technologies: An empirical analysis in Italy," *Sustain. Energy Technol. Assessments*, vol. 15, pp. 1–10, Jun. 2016.
[51] H. Dunkelberg *et al.*, "Optimization of the energy supply in the plastics industry to reduce the primary energy demand," *J. Clean. Prod.*, vol. 192, pp. 790–800, Aug. 2018.
[52] A. Beghi, M. Lionello, and M. Rampazzo, "Energy-efficient management of a wood industry facility," in *2017 IEEE Conference on Control Technology and Applications (CCTA)*, 2017, pp. 181–186.
[53] S. Karellas, D. Giannakopoulos, C.-S. Hatzilau, I. Dolianitis, G. Skarpetis, and T. Zitounis, "The potential of WHR/batch and cullet preheating for energy efficiency in the EU ETS glass industry and the related energy incentives," *Energy Effic.*, vol. 11, no. 5, pp. 1161–1175, Jun. 2018.
[54] S. Seidel, M. Franke, F. B. H. Wilson, and U. Gromnitza, "Enhancing the energy-efficient production of tempered glass by using simulation-based optimisation," in *2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2017, pp. 1–4.
[55] R. M. Lazzarin and M. Noro, "Energy efficiency opportunities in the production process of cast iron foundries: An experience in Italy," *Appl. Therm. Eng.*, vol. 90, pp. 509–520, Nov. 2015.
[56] M. Noro and R. M. Lazzarin, "Energy audit experiences in foundries," *Int. J. Energy Environ. Eng.*, vol. 7, no. 4, pp. 409–423, Dec. 2016.
[57] R. M. Lazzarin and M. Noro, "Energy efficiency opportunities in the service plants of cast iron foundries in Italy," *Int. J. Low-Carbon Technol.*, May 2016.
[58] M. Sun, Y. Wang, L. Shi, and J. J. Klemeš, "Uncovering energy use, carbon emissions and environmental burdens of pulp and paper industry: A systematic review and meta-analysis," *Renew. Sustain. Energy Rev.*, vol. 92, pp. 823–833, Sep. 2018.
[59] P. W. Griffin, G. P. Hammond, and J. B. Norman, "Industrial decarbonisation of the pulp and paper sector: A UK perspective," *Appl. Therm. Eng.*, vol. 134, pp. 152–162, Apr. 2018.

[60] H. Holmberg, M. Tuomaala, T. Haikonen, and P. Ahtila, “Allocation of fuel costs and CO2-emissions to heat and power in an industrial CHP plant: Case integrated pulp and paper mill,” *Appl. Energy*, vol. 93, pp. 614–623, May 2012.
[61] K. Möllersten, L. Gao, J. Yan, and M. Obersteiner, “Efficient energy systems with CO2 capture and storage from renewable biomass in pulp and paper mills,” *Renew. Energy*, vol. 29, no. 9, pp. 1583–1598, Jul. 2004.
[62] P. Ahmadi, A. Almasi, M. Shahriyari, and I. Dincer, “Multi-objective optimization of a combined heat and power (CHP) system for heating purpose in a paper mill using evolutionary algorithm,” *Int. J. Energy Res.*, vol. 36, no. 1, pp. 46–63, Jan. 2012.
[63] D. J. Marshman, T. Chmelyk, M. S. Sidhu, R. B. Gopaluni, and G. A. Dumont, “Energy optimization in a pulp and paper mill cogeneration facility,” *Appl. Energy*, vol. 87, no. 11, pp. 3514–3525, Nov. 2010.
[64] Eurostat, “Electricity cost data (Accessed 06/2018).” [Online]. Available: http://appsso.eurostat.ec.europa.eu/nui/show.do. [Accessed: 21-Jun-2018].
[65] Eurostat, “Natural Gas cost Data (Accessed 06/2018).” [Online]. Available: http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=nrg_pc_203&lang=en. [Accessed: 21-Jun-2018].
[66] B. B. C. J. Manyika, “Are you ready for the era of ‘big data’?,” *Mc Kinsey*, 2011. [Online]. Available: https://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/are-you-ready-for-the-era-of-big-data. [Accessed: 05-Aug-2019].
[67] Viktor Mayer-Schönberger; Kenneth Cukier, *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt, 2013.
[68] K. Zhou, C. Fu, and S. Yang, “Big data driven smart energy management: From big data to big insights,” *Renew. Sustain. Energy Rev.*, vol. 56, pp. 215–225, Apr. 2016.
[69] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY: Springer New York, 2009.
[70] R Pubs, “Iris Dataset - Clustering using K means.” [Online]. Available: https://rpubs.com/AnanyaDu/361293. [Accessed: 10-Aug-2019].
[71] A. Hyvärinen, “A unified probabilistic model for independent and principal component analysis,” in *Advances in Independent Component Analysis and Learning Machines*, Elsevier, 2015, pp. 75–82.
[72] Nasruddin, N. Aisyah, M. I. Alhamid, B. B. Saha, S. Sholahudin, and A. Lubis, “Solar absorption chiller performance prediction based on the selection of principal component analysis,” *Case Stud. Therm. Eng.*, vol. 13, p. 100391, Mar. 2019.
[73] Wikipedia, “File:Manual decision tree.jpg.” [Online]. Available: https://en.wikipedia.org/wiki/File:Manual_decision_tree.jpg. [Accessed: 10-Aug-2019].
[74] M. W. Ahmad, M. Mourshed, and Y. Rezgui, “Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption,” *Energy Build.*, vol. 147, pp. 77–89, Jul. 2017.
[75] Ü. Ç. Büyükşahin and Ş. Ertekin, “Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition,” *Neurocomputing*, vol. 361, pp. 151–163, Oct. 2019.
[76] F. Martinez Alvarez, A. Troncoso, J. C. Riquelme, and J. S. Aguilar Ruiz, “Energy Time Series Forecasting Based on Pattern Sequence Similarity,” *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 8, pp. 1230–1243, Aug. 2011.
[77] R. Talavera-Llames, R. Pérez-Chacón, A. Troncoso, and F. Martínez-Álvarez, “Big data time series forecasting based on nearest neighbours distributed computing with Spark,” *Knowledge-Based Syst.*, vol. 161, pp. 12–25, Dec. 2018.
[78] R. Talavera-Llames, R. Pérez-Chacón, A. Troncoso, and F. Martínez-Álvarez, “MV-kWNN: A novel multivariate and multi-output weighted nearest neighbours algorithm for big data time

series forecasting," *Neurocomputing*, vol. 353, pp. 56–73, Aug. 2019.
[79] B. Yuce, Y. Rezgui, and M. Mourshed, "ANN–GA smart appliance scheduling for optimised energy management in the domestic sector," *Energy Build.*, vol. 111, pp. 311–325, Jan. 2016.
[80] I. P. Panapakidis, T. A. Papadopoulos, G. C. Christoforidis, and G. K. Papagiannis, "Pattern recognition algorithms for electricity load curve analysis of buildings," *Energy Build.*, vol. 73, pp. 137–145, Apr. 2014.
[81] Y. Kurtgoz, M. Karagoz, and E. Deniz, "Biogas engine performance estimation using ANN," *Eng. Sci. Technol. an Int. J.*, vol. 20, no. 6, pp. 1563–1570, Dec. 2017.
[82] H. K. Ghritlahre and R. K. Prasad, "Application of ANN technique to predict the performance of solar collector systems - A review," *Renew. Sustain. Energy Rev.*, vol. 84, pp. 75–88, Mar. 2018.
[83] A. Tugcu and O. Arslan, "Optimization of geothermal energy aided absorption refrigeration system—GAARS: A novel ANN-based approach," *Geothermics*, vol. 65, pp. 210–221, Jan. 2017.
[84] L. Zhou *et al.*, "Analysis of influencing factors of the production performance of an enhanced geothermal system (EGS) with numerical simulation and artificial neural network (ANN)," *Energy Build.*, vol. 200, pp. 31–46, Oct. 2019.
[85] V. Z. Antonopoulos, D. M. Papamichail, V. G. Aschonitis, and A. V. Antonopoulos, "Solar radiation estimation methods using ANN and empirical models," *Comput. Electron. Agric.*, vol. 160, pp. 160–167, May 2019.
[86] C. Voyant *et al.*, "Machine learning methods for solar radiation forecasting: A review," *Renew. Energy*, vol. 105, pp. 569–582, May 2017.
[87] S. Sobri, S. Koohi-Kamali, and N. A. Rahim, "Solar photovoltaic generation forecasting methods: A review," *Energy Convers. Manag.*, vol. 156, pp. 459–497, Jan. 2018.
[88] U. K. Das *et al.*, "Forecasting of photovoltaic power generation and model optimization: A review," *Renew. Sustain. Energy Rev.*, vol. 81, pp. 912–928, Jan. 2018.
[89] H. Liu, C. Chen, X. Lv, X. Wu, and M. Liu, "Deterministic wind energy forecasting: A review of intelligent predictors and auxiliary methods," *Energy Convers. Manag.*, vol. 195, pp. 328–345, Sep. 2019.
[90] MATLAB, "K-Mean function - MATLAB." [Online]. Available: https://it.mathworks.com/help/stats/kmeans.html. [Accessed: 19-Jan-2019].
[91] J. E. Gentle, L. Kaufman, and P. J. Rousseuw, *Finding Groups in Data*, vol. 47, no. 2. Hoboken, NJ, USA: John Wiley & Sons, Inc., 1990.
[92] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987.
[93] MATLAB, "FitchkNN function - MATLAB." [Online]. Available: https://it.mathworks.com/help/stats/fitcknn.html. [Accessed: 23-Jun-2019].
[94] A. Çay, "Energy consumption and energy saving potential in clothing industry," *Energy*, vol. 159, pp. 74–85, Sep. 2018.
[95] E. Andersson, M. Karlsson, P. Thollander, and S. Paramonova, "Energy end-use and efficiency potentials among Swedish industrial small and medium-sized enterprises – A dataset analysis from the national energy audit program," *Renew. Sustain. Energy Rev.*, vol. 93, pp. 165–177, Oct. 2018.
[96] V. Ghalandari, M. M. Majd, and A. Golestanian, "Energy audit for pyro-processing unit of a new generation cement plant and feasibility study for recovering waste heat: A case study," *Energy*, vol. 173, pp. 833–843, Apr. 2019.
[97] A. Biglia, F. V. Caredda, E. Fabrizio, M. Filippi, and N. Mandas, "Technical-economic feasibility of CHP systems in large hospitals through the Energy Hub method: The case of Cagliari AOB," *Energy Build.*, vol. 147, pp. 101–112, Jul. 2017.
[98] Capstone Turbine Corporation, "Capstone Model C200 Performance." [Online]. Available:

http://www.regattasp.com/files/410066C_C200_Tech_Ref.pdf. [Accessed: 19-Jan-2019].
[99] M. A. R. Biswas, M. D. Robinson, and N. Fumo, "Prediction of residential building energy consumption: A neural network approach," *Energy*, vol. 117, pp. 84–92, Dec. 2016.
[100] D. Koschwitz, J. Frisch, and C. van Treeck, "Data-driven heating and cooling load predictions for non-residential buildings based on support vector machine regression and NARX Recurrent Neural Network: A comparative study on district scale," *Energy*, vol. 165, pp. 134–142, Dec. 2018.
[101] K. Cheng, L. M. Guo, Y. K. Wang, and M. T. Zafar, "Application of clustering analysis in the prediction of photovoltaic power generation based on neural network," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 93, p. 012024, Nov. 2017.
[102] D. Yang, Z. Dong, L. H. I. Lim, and L. Liu, "Analyzing big time series data in solar engineering using features and PCA," *Sol. Energy*, vol. 153, pp. 317–328, Sep. 2017.
[103] M. Malvoni, M. G. De Giorgi, and P. M. Congedo, "Photovoltaic forecast based on hybrid PCA–LSSVM using dimensionality reducted data," *Neurocomputing*, vol. 211, pp. 72–83, Oct. 2016.
[104] M. Malvoni, M. G. De Giorgi, and P. M. Congedo, "Data on Support Vector Machines (SVM) model to forecast photovoltaic power," *Data Br.*, vol. 9, pp. 13–16, Dec. 2016.
[105] S. Qijun, L. Fen, Q. Jialin, Z. Jinbin, and C. Zhenghong, "Photovoltaic power prediction based on principal component analysis and Support Vector Machine," in *2016 IEEE Innovative Smart Grid Technologies - Asia (ISGT-Asia)*, 2016, pp. 815–820.
[106] Y. Amri, A. L. Fadhilah, Fatmawati, N. Setiani, and S. Rani, "Analysis Clustering of Electricity Usage Profile Using K-Means Algorithm," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 105, p. 012020, Jan. 2016.
[107] H. Müller, "CLASSIFICATION OF DAILY LOAD CURVES BY CLUSTER ANALYSIS," in *Proceedings of the Eighth Power Systems Computation Conference*, Elsevier, 1984, pp. 381–388.
[108] F. Wahid and D. Kim, "A Prediction Approach for Demand Analysis of Energy Consumption Using K-Nearest Neighbor in Residential Buildings," *Int. J. Smart Home*, vol. 10, no. 2, pp. 97–108, Feb. 2016.
[109] "A Novel Hybrid Model Based on Extreme Learning Machine, k-Nearest Neighbor Regression and Wavelet Denoising Applied to Short-Term Electric Load Forecasting," *Energies*, vol. 10, no. 5, p. 694, May 2017.
[110] B. Lantz, *Machine Learning with R*, October 20. Packt Publishing.