End of Master's Degree Project

DECORRELATION BETWEEN SOLAR AND WIND POWER AND ITS EFFECTS ON ENERGY STORAGE DIMENSIONING

Memory

Author: Luis Barrantes Coloma
Director: Vicente Cesar de Medina Iglesias

June 2021

Superior Technical School of Industrial
Engineering of Barcelona

ETSEIB

UPC

Abstract

The energetic generation is a widely spread topic due to the great environmental impact it makes, the high value of resources it consumes and the controversy it creates; not only that, but the great infighting between renewable and non-renewable energy sources which started years ago and doesn't seem to be ending anytime soon, may be the biggest discussion topic in that field.

This project will study the effect of the combination of solar photovoltaic and wind power, two of the most well known renewable energies, and the effect they have when covering the energy demand of a country, in this case Spain.

To achieve this objective, several combinations of places from the Iberian Peninsula will be used along other overseas locations as reference. From these places a qualitative study has been done first, where we obtain that the places whose sun and wind values are higher and, at the same time cover their own deficits; are the best locations to cover the energy demand.

After that, a second study has been made, where a fictitious storage was created for the same combinations as in the first part; the main goal of this part was to find out the best possible combination of places that gave the minimum storage with the least energy deficit time. The optimal values were found to be, again, those with the best solar and wind decorrelation.

Index

1. Foreword

1.1. Origin of the project

Nowadays, the energetic consumption is not only high, but increases as each day passes; not only that, but the energy obtaining methods are still the same that were used fifty years ago, so the polluting emissions are not only not dwindling, but increasing year by year.

To try to counter that, the renewable energies have, in the recent years, crashed into the energy generation mix, even though they still don't have the strength to compete versus the classic generations. The most clear examples are the photovoltaic sun and wind energy which, despite having an infinitely higher potential that the conventional energy sources, have also vey important drawbacks.

This is the reason that the idea of combining together these renewable energy sources came to life so that they could, by working together, cover their negative points and turn into an energy generation powerhouse and therefore, this project was born.

1.2. Motivation

The main motivation of this project is to prove that solar and wind power can work together as a stable base energy generation; similar to the nuclear power generation nowadays. This is due to the fact that if we were to obtain that kind of energy generation we could, along with hydro generation to cover demand spikes, obtain an energetic mix only using renewables energies.

It is an obvious statement to say though, that the renewable technology we have nowadays it's not on the same level that the rest of energy sources like coal, combined cycle or nuclear; this statement is especially true for the solar photovoltaic generation. Having said that, new renewable energy research is being carried out day by day and thus, the renewable energies are slowly but steadily gaining terrain on the non renewable ones.

This is especially true for the case of Spain, due to the fact that it has a huge renewable potential; it has a lot of different places where hydro power can be installed, a lot of strong winds where wind farms can be installed and most importantly, a degree of solar radiation higher than any other point on Europe. This is why, this new method of solar and wind power compensation has such a high potential and thus, has to be explored so we can make of this country and world a greener place.

2. Introduction

2.1. Project objectives

This project has as objective to study the relation between solar photovoltaics and wind power generation and how can they work together to compliment their inherent unreliability when it comes to their energetic production.

This project will consist in two main blocks; both with a theoretical part, where we will study important concepts that we will use in their respective practical sections. These practical sections will consist in the following:

The first section will be a qualitative study of different locations in Spain and some other overseas places, in here we will see how does the power generation obtained by combining different places adapts to the demand.

The second part will be a more quantitative study of the previous part; in here we will calculate a storage value and installed power coefficients for each solar and wind generation and choose the ideal combination and values for the adaptation of a certain demand.

2.2. Project scope

This project's scope will be the following:

- Study of the different sources of energy that are used nowadays to generate electricity.

- Creation of a Python program that allows us to study the different locations' sun and wind parameters and the results obtained after combining different places.

- Compare how well the sun and wind from different locations combined cover the demand from Spain.

- Study of the different types of energy storage and their main characteristics.

- Creation of a second Python program that allows us to create a fictitious storage system for each different solar and wind combination and the installed power values needed to cover that storage.

- Compare the different storage and coefficients and choose the ideal location between the studied data sets.

3. The energy generation situation nowadays

We will begin this project by talking about the energy situation in today's world. As we know, we humans are in a state of constant growth, this implies also a constant energy demand increase year after year, therefore, a huge investment in its generation has to be done constantly, ever increasing as years pass.

In this section we will be talking about these sources of energy and their distribution on the world, explain the main points of the most common ones and introduce the concept of decorrelation, which will be the base of our entire project.

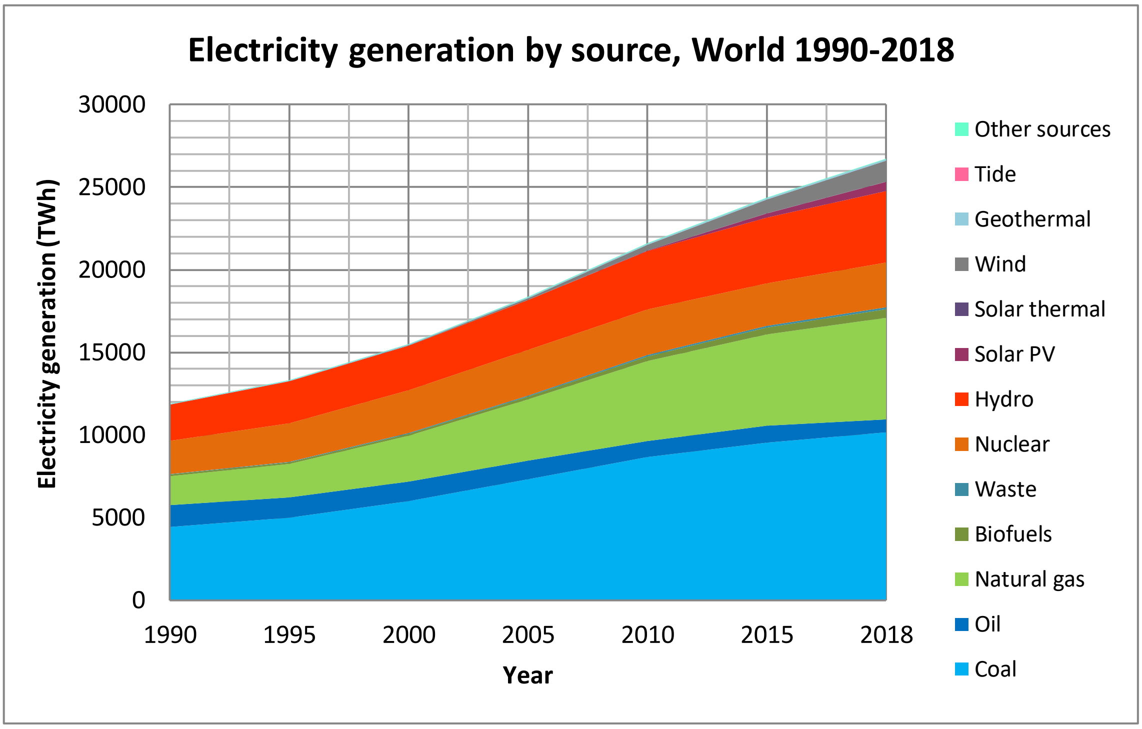
3.1. Energy distribution around the globe

There are a lot of different ways if obtaining energy, from simply burning combustible and extracting their thermal energy to other more complicated methods like the transformation of natural phenomena into energy (like sun, wind or tides) and everything between them.

This energy, obtained through very different methods, has many uses; from the transformation of products in industrial sectors to the heating of homes. But the most important one between all of them is the electricity generation due to the fact that not only electricity is used for lighting or heating, but it is also used to power almost everything us humans use, like transport, industry or even the same energy extraction, which sometimes requires the use of electricity, either directly or indirectly.

Therefore, the vast majority of all the energy sources in the world, with the exception of some gas products and oil, which have more important applications in the industrial and transportation sector and bio-waste, which is mostly used in heating; are used in the electricity production as we can see in the Sankey diagram of the world energy balance in 2018, which due to space constraints will be shown in the Annex 2: Sankey diagram.

As we can see, a lot is invested in electricity generation, increasingly with every year; in the following *Plot 3-1* we can also see the world's electricity generation by source from 1990 to 2018:



Plot 3-1: Electricity generation by source; World 1990-2018

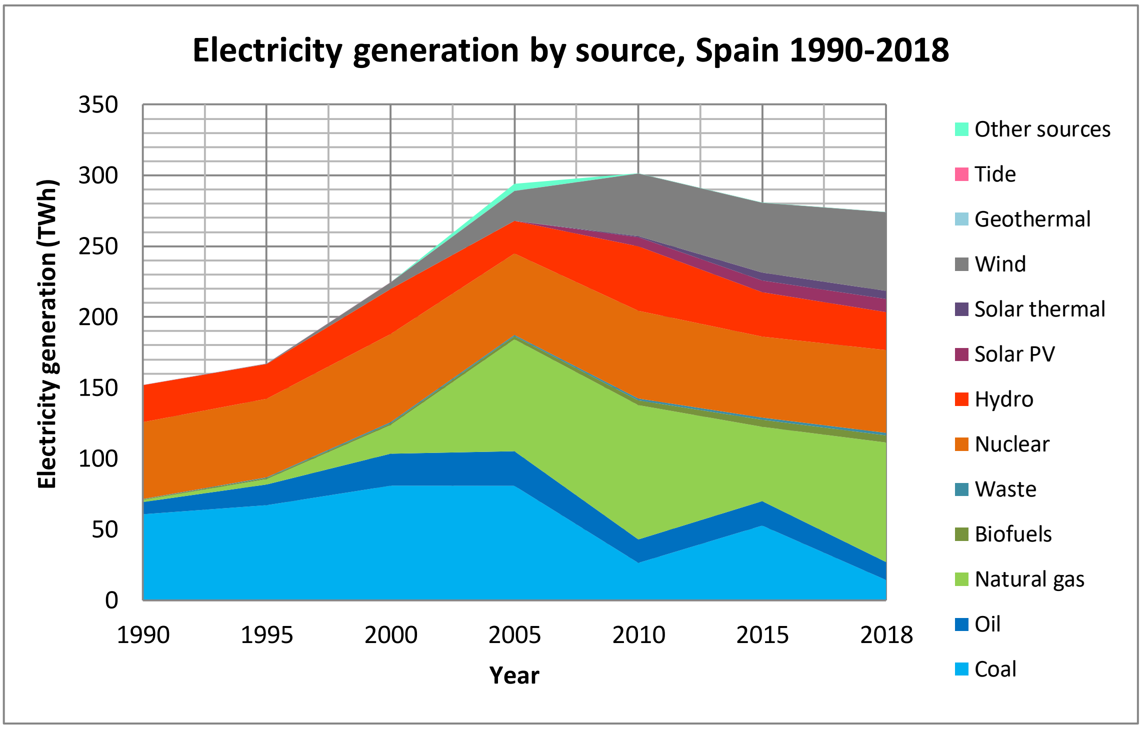
As we may appreciate, all energy generations increase as every year passes; the most important being coal, natural gas, hydro electrical power and nuclear. It is also important to mention that, as the plot shows, more than half of the world electricity is produced by non renewable sources, with the coal being the most used one. This of course represents a problem in the long run due to the fact of them being, as their name says, non-renewable, even though this doesn't seem to stops their increase of use every year, despite the increase of heavy restrictions that some governments keep applying on non-renewable combustibles.

From these top four electricity sources mentioned before, the only renewable energy capable of competing with the other non-renewable ones is the hydro power due to the amount of power it can yield, its high efficiency and the abundance of lakes and rivers all along Earth.

We can also see than the other known renewable energies (mainly solar PV and wind), which appeared around 2005 are starting to gain more popularity every year it passes; having said that, their generation value is nowhere near the magnitude of the main four generations.

3.1.1. Energy distribution in Spain

In this project, we will be focusing on various locations from Spain, and some other overseas places to use as reference, so it can be useful to know beforehand how is the overall electricity situation in the country; therefore, in the following *Plot 3-2* we will be able to see the electricity generation mix divided by source:



Plot 3-2: Electricity generation by source; Spain 1990-2019

When looking at the Spanish electricity generation plot we can see two important differences when comparing to the overall world generation.

The first important thing to notice is the fact that not only Spain has a relative low coal use value compared to the rest of the world, but also it keeps decreasing every year; mostly being replaced by wind power and natural gas. This is due to the fact that the Spanish government for the recent years has been pushing the closure of all coal power plants all over the country [1].

The reason that the increase of the use of natural gas is due to the decrease in its price in the recent years, which the combined cycle power plants use mainly as fuel.

If we look at the renewable aspect of the graph, we can see that the Spanish values are relatively good, especially in the wind category, with almost another 20% of the total electricity generation. As for the rest of the renewable mix, we can see that Spain also has a consistent hydro generation and a small, but increasing, solar PV production.

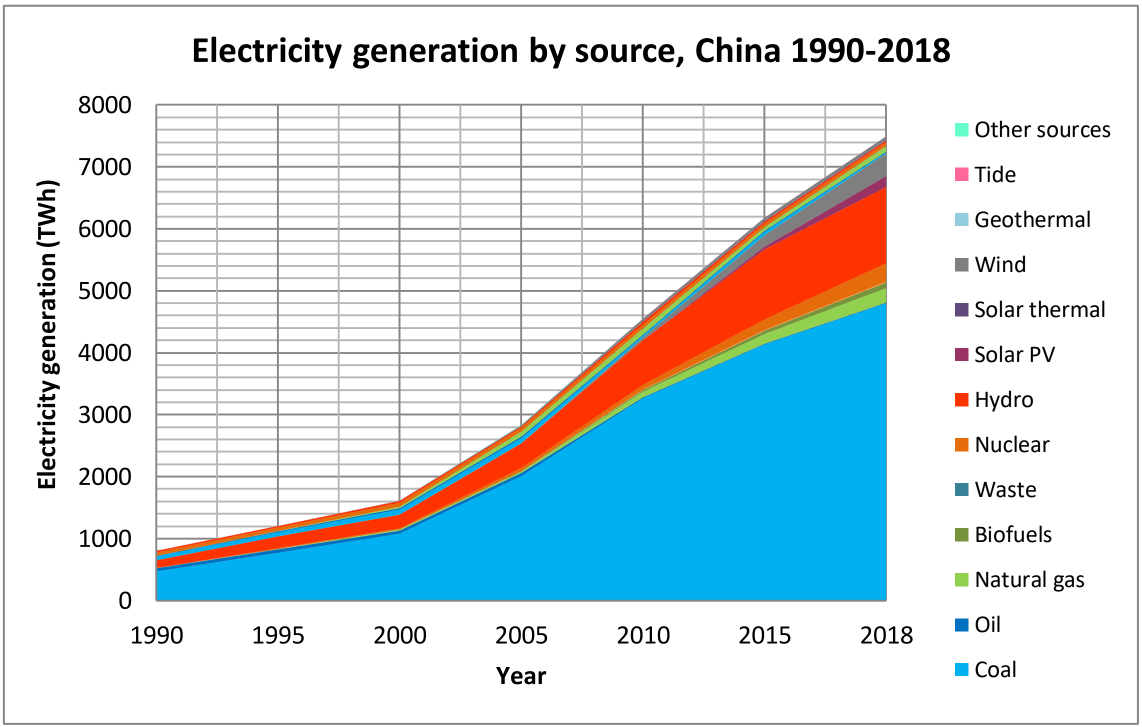
Finally, Spain also has a relatively high and stable nuclear energy generation, despite the fact of only having 7 operating power plants all around the country [3], which where built in the 80s, meaning they are relatively old, but they still generate around the 20% of the total electricity in the country.

Overall we can see that Spain is trying very hard to reduce the amount of combustible fuel used in the energy generation and trying to replace it with cleaner and greener energy sources.

3.2. Main concerns about energy generation

Nowadays there is a very big concern with global warming and CO2 emissions all over the world, therefore a lot of restrictions are being imposed on countries to their energy generation and use of combustible fuels like coal or natural gas so the polluting emissions to the atmosphere lower.

Even though there are heavy restrictions applied, the use of coal and other polluting energy sources keeps increasing every year, as we can see in the following *Plot 3-3* of China's electricity production:



Plot 3-3: Electricity generation by source; China 1990-2019

As we can clearly see, and if we compare it with the world global generation we saw in *Plot 3-1*; we can see that China consumes more that 50% of the coal in the world [4]; therefore the harsh restrictions that may be applied to other smaller countries may not have much of an effect if there is such a huge consumption that undoes what others do.

This, though, is a difficult thing to change, due to the fact of it being highly dependent of every country and its legislations. Therefore, there is the need to contribute with as much as possible to the reduction of such polluting effects.

This is why, since several years, a lot of countries, Spain included, have been advocating and implementing several renewable energy methods in the energy mix while, at the same time, reducing the polluting energy sources.

Having said that, renewable energies are not perfect, because they have flaws, like the non-renewable energies. In the following point we will be explaining some of the weaknesses of each energy generation.

3.2.1. Fossil fuel energy

This category includes the electricity generated by the use of coal, natural gas and oil (even though this last one is not that widely used in electricity production as seen in the Sankey diagram on *Annex 2: Sankey diagram*. What these types of energy have in common are that they use the burning of the combustible to generate heat, which eventually will turn into electricity. The main problem with this process is that the combustion of these materials, which are composed mainly from hydrocarbons and some other impurities line sulfur (S) or nitrogen (N) emit, when burned, polluting gases like carbon dioxide (CO2), carbon monoxide (CO), sulfur dioxide (SO2) or nitrogen dioxide (NO2) which escape to the atmosphere and produce not only greenhouse effect, but also phenomena like acid rain.

This is the main reason that governments are trying to reduce the use of these technologies and replacing them with other greener alternatives.

3.2.3. Nuclear energy

Nuclear energy is also a commonly used energy all over the world, maybe not as much as other big generations, but with relatively high values overall. The main strength that nuclear energy has compared to any other energy generation source is the vast quantity of energy it can produce while using a very small amount of fuel. Also the emissions, if we don’t count the spent combustible are completely clean from polluting components, in contrast of the fossil fuel sources.

The negative points of this energy are the ones that heavily weight it down, due to the fact that the waste it generates is radioactive, meaning that its effect linger over long periods of time, therefore, if not contained properly, can create severe damage to the environment.

Another very strong point against the nuclear energies is the fact that when accidents occur, even though nuclear power plants heavily invest in security through different layers of protection, the results are very catastrophic for all life and environment; as happened in the most well known nuclear accidents in history, Chernobyl, Three Mile Island of Fukushima.

Therefore, it doesn't come as a surprise that some countries are reducing the use of nuclear power plants (like Germany) or straight away banning it for energetic purposes (like Italy) [5]. Having said that there are countries heavily invested in nuclear energy, like France or the USA.

3.2.2. Hydro electrical energy

Hydro electrical energy is the most used renewable energy source in the world; this is due to the fact that it is able to generate huge quantities of power at a high efficiency without any emission of greenhouse or polluting gasses.

However, it is not free of flaws [6], due to the fact that despite not polluting, it still causes environmental impact because its construction can damage local wildlife habitats, their migratory paths and the overall ecosystem, damage the quality of the water and alter rivers and lakes.

It also requires a big terraforming, both to create the reservoirs, if they are not natural; and to the installation of the damns, which can't be installed everywhere. This terraforming contributes significantly to the increase of the greenhouse effect.

Despite this, hydro power is a very viable type of energy overall, even more if compared with the non-renewable ones.

3.2.4. Solar and wind energy

Finally we get at the two rising renewables; the solar (mostly photovoltaic) and wind. These two type of energies, as mentioned before, are increasing year after year in the generation mix, due to their clean generation and huge potential more countries are increasing the budget dedicated to them, having said that, these two energy sources have each one their own problems.

The solar photovoltaic has the problem of having a very low efficiency when compared to other types of generation, like for example hydro or the combustion ones; it also has a very high installation cost and storage, due to generating electricity directly which is way more difficult to store than solid fuel or water.

Then we have the wind power, which main disadvantages consist of noise pollution, a high cost inversion and visual impact; also the space requirement for this power source is very big, around 0.3 hectares per MW [7]; even though the land between turbines can be used for other purposes. Another negative point is the fact that they are limited by the wind speed; meaning they can't work at very low or very high speeds.

But if we had to choose only one negative point, the main disadvantage for both solar and wind energies is their unreliability, due to the fact that the photovoltaics only produce energy on the day and not only that, but clouds also heavily reduce their efficiency. As for the wind, it can blow either at day or night, but it can’t be predicted quite easily and is very inconsistent.

This is, of course a very important disadvantage, because if those energies were constant and reliable, their values would skyrocket due to their infinite supply and zero emissions.

Here is where the goal of this project comes to life; we know that sun and wind by themselves are both unreliable, but as we also said, wind can blow day or night, so if we were able to combine sun and wind energy to compensate between them when one becomes weaker; we would obtain a very strong, decorrelated generation combo which could eventually outmatch other energy sources.

Before talking about this decorrelation concept we will proceed to do a brief explanation of what are and how these energy sources do work, so we get to know them better.

3.3. Solar power

As everyone may know, solar power consists on the use of the Sun radiation to obtain usable energy, let it be for heating, water treatment or electricity generation [8].

Solar power is obtained through the irradiance that the Sun emits and arrives to Earth; this radiation will be different depending on the location of the Earth; it will be higher in the places closer to the tropics and lower in the poles. In the *Image 3-4* we can see the world solar energy map and the annual average irradiance:

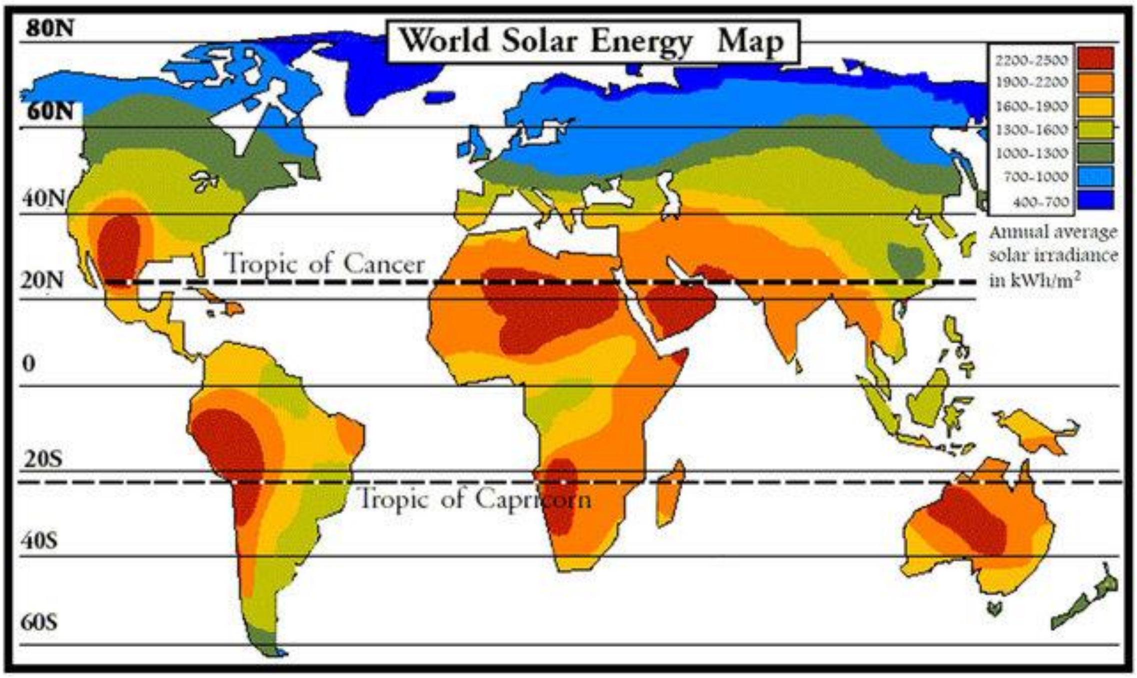


Image 3-4: World solar energy map

As we can see, the solar irradiance focuses in the Tropic of Cancer and Capricorn, due to the fact of them being the zones closer to the Sun on average during the year.

If we look at the situation in Spain we can see that the average irradiation is between 1300 and 1900 kWh/m^2 an average value compared with the rest of the world.

As we mentioned before, solar power can be used in plenty of different ways; in the following points we will briefly explain the most common ways of using this power:

3.3.1. Passive solar energy

The use of passive solar energy consists in the use of the solar radiation without any transformation at all; it is used by taking into account the design and placement of the different building components to make use of the sunlight for daylight, space heating of houses and greenhouses or if used correctly, even cooling [8]. These systems are characterized by having a very low installation and maintenance cost.

3.3.2. Thermal solar energy

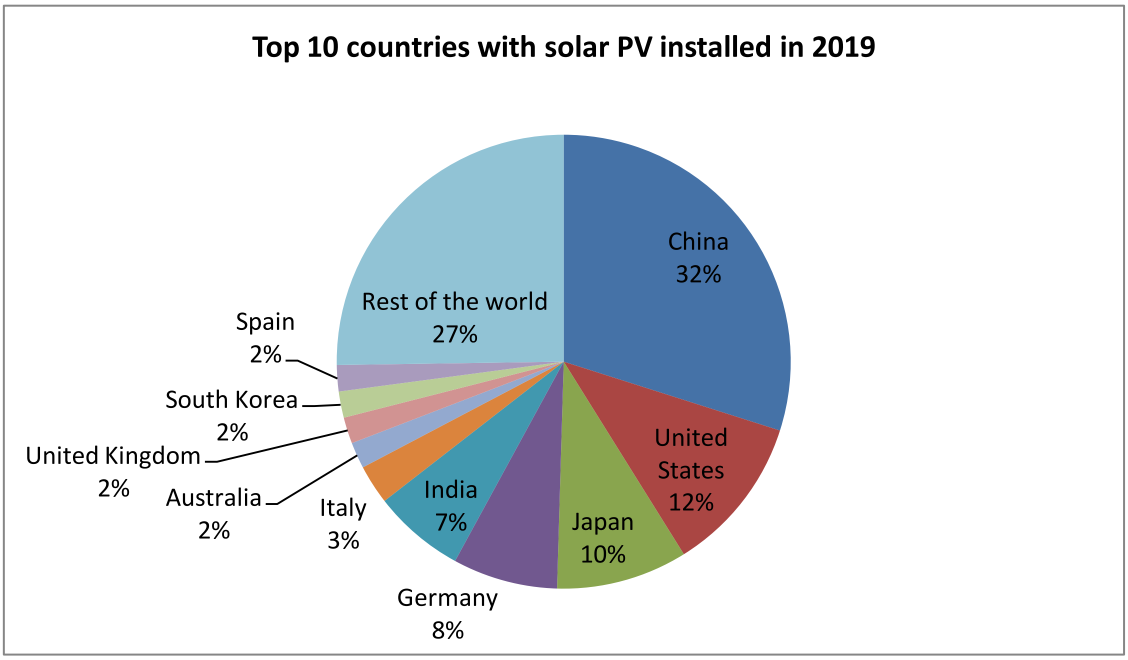
This method consists on the use of the Sun's innate heating power to transfer it to a heating transfer medium (like water, air or molten salts), after that this heating source will be send to generate electricity by similar means as the thermal power centrals.

All the solar thermal plants have two main components; reflectors, that captures and reflects the sunlight and a receiver, which captures that light and heats the transfer medium.

3.3.3. Photovoltaic (PV) solar energy

We arrive now at the most popular solar energy generation, due to the fact that it can convert directly the sun radiation into electrical energy by using the photovoltaic effect and specially designed solar panels, and thus, avoiding energy transformation which all have inherent losses; having said that, solar PV energy has a relative low efficiency (around 20% for the best panels in the market [9]). Originally, this type of energy was created to provide energy to the satellites at outer space, but today the vast majority is used in grid-connected systems for power generation.

The vast majority of countries around the world are, nowadays heavily increasing the funding for solar PV energy due to being non-polluting and huge potential development. In the following *Plot 3-5* we can see the top 10 countries with solar photovoltaic energy installed.



Plot 3-5: Top 10 countries with solar PV installed in 2019

As we can see, the country which has the most solar PV installed is China, with almost 1/3 of the global production; even though if we look at the total electricity generation from China in *Plot 3-3* we can see that it is still a small quantity. Having said that, as we also saw in the global electricity generation in *Plot 3-1* the percentage of the solar PV is steadily increasing every year.

As for the other countries we can see that along with China, Japan and the USA have almost the 50% of all the solar generation in the world.

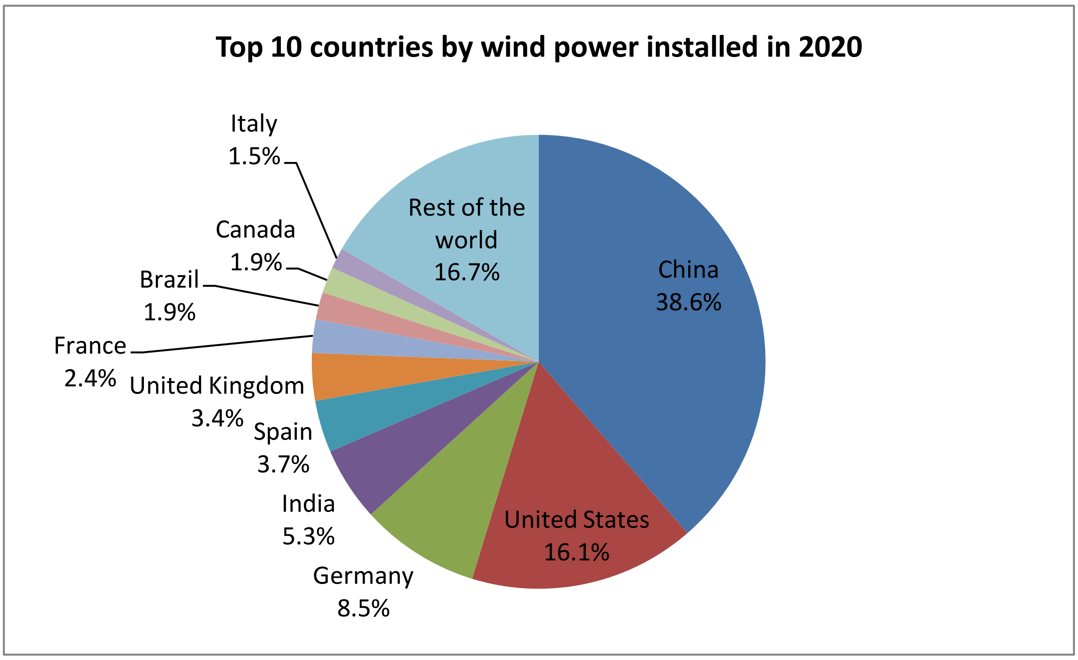
In the case of Spain we can see that it occupies the 10th position in photovoltaics generation, which is a very high value. This has been possible thanks to the heavy investment that it has been done in promoting and installing solar power plants and closing coal power plants in the recent years.

With this we now have a wider view of the solar power generation, and more specifically the photovoltaics one; therefore, the next step will be to do a similar study of the wind power, which will be done in the next section.

3.4. Wind power

Wind power consists in the use of the wind's own kinetic energy to provide mechanical power through wind turbines and use it to create electrical power using electric generators. This energy source is very clean and a relative high efficiency one.

As with the solar power, the wind power has been gaining popularity among the energy sources with a steady increase in total installed power year after year, but as we said before, it still not in the same level as the other major sources. In the following *Plot 3-6* we can see the top 10 countries with wind power installed:



Plot 3-6: Top 10 countries with wind power installed in 2020

We can see that, once again China is the leading country in wind installed power with more than a third of the total installed capacity, which again compared with its total generation is pretty small.

As for the rest of countries we can see that China is followed by the USA, India, and Germany and after all those Spain with a 3.7% of the world total. As we saw when we talked about the energy generation mix in Spain, the wind power generation is one of the most important ones, with an increasing value every year.

3.4.1. Wind power generation

The wind power is defined as the wind's kinetic energy per unit of time that goes through the wind generator's rotor area, shown in the following equation [10]:

$$P(v) = \frac{1}{2}A\rho v^3$$

This previous equation doesn't directly tell us the amount of electrical power generated by a wind turbine though; because a lot of other parameters have to be taken into account, like the turbine efficiency or the blade design.

One important thing to mention is that wind turbines can't work at any desired wind speed; if it is too low it's own inertia and friction will make it impossible to move (cut-in speed), if the wind speed is too high there is risk of breaking the turbine and it has to be stopped (cut-out speed). Also there is a certain speed, at which the turbine power is capped known as the rated speed. In the following *Image 3-7* we can see the representation of said curve for a typical wind turbine.

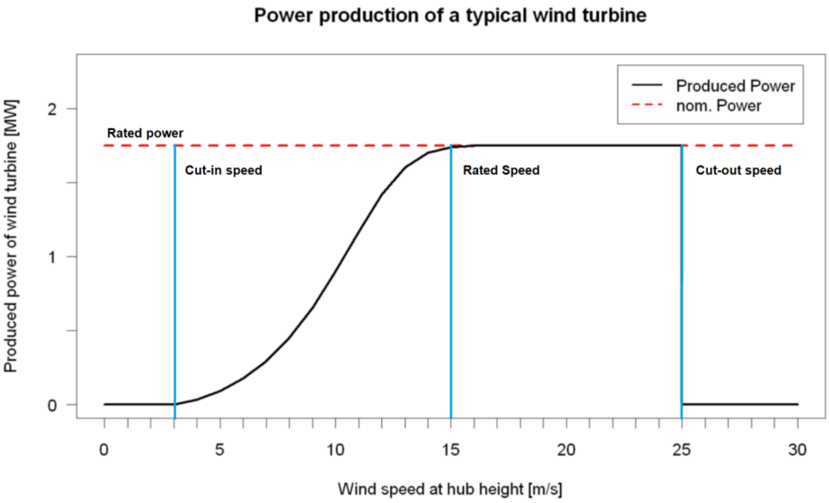
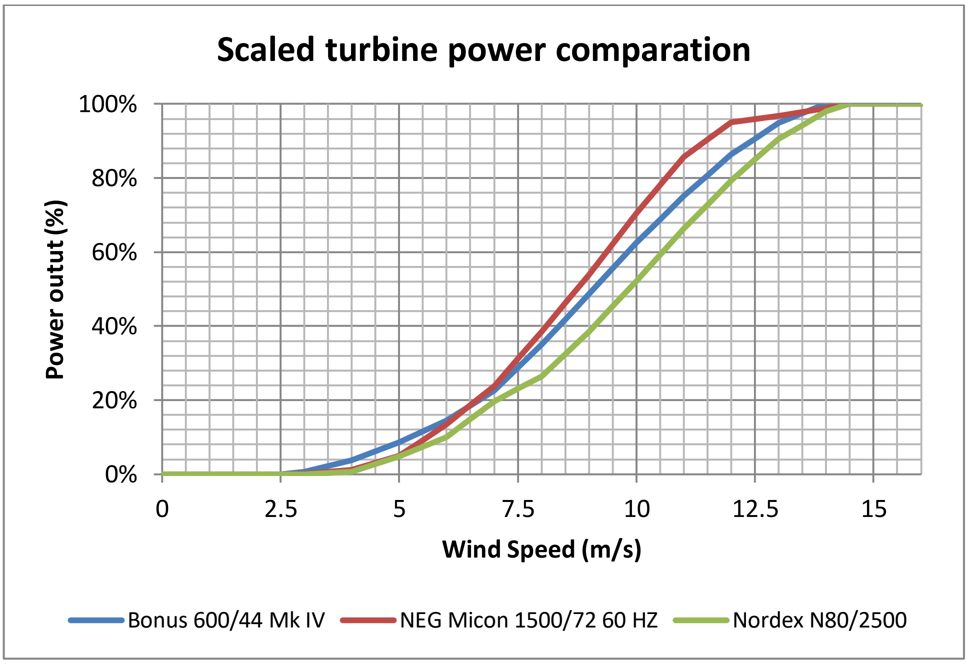


Image 3-7: Power production curve of a typical wind turbine

Most power generators follow this previous curve, with values of 0 for speeds below 3 m/s and over 25 m/s; a maximum power output determined by their installed capacity between 15 and 25 m/s and an increasing sigmoid curve between 3 and 15 m/s. This power curve may not always give us the exact value of power generated because the irregularity of the wind velocities may make it vary, but their value is sufficiently steady to be used in this project. The magnitudes of said curve may differ from each other, but their profile doesn't. This is why if we standardize all the turbines power curves we will obtain similar power curves.

In the following plot we can see the power curve of three different wind turbines: Bonus 600/44 Mk IV [11]; NEG Micon 1500/72 60 HZ [12] and Nordex N80/2500 [13]. In this case, because all of them have different power output, they have been scaled by dividing by the maximum value; the results can be seen in the *Plot 3-8*:



Plot 3-8: Scaled turbine power comparison

As we can see, all three power curves have very similar curves which will be very useful when trying to calculate the wind values in the practical part because we will be able to make a linear approximation to said curves. One important thing to mention is the fact that in this previous plot we have only shown the variation between cut in speed and rated speed; this is due to the fact that, even though sometimes the wind turbine may have small losses during their rated speed section, we will consider them non existent in order to simplify the calculations of the practical study; which will be expanded in further sections of the project.

3.5. Decorrelation energy concept

After having studied the solar and wind generations we will introduce the concept that will be the base of the project's study; this is the concept of decorrelation between solar and wind generation.

As mentioned in previous sections, sun and wind energy are very unreliable; while the sun only produces power in day hours (approximately between 6 am and 6 pm, depending on the location); the wind can blow at any hour, but it has not fixed values or patterns. Therefore, their individual use is one that can not be considered as a stable generation source. But then a question arises:

Even though both of them are unreliable, if we combine them, won't they be able to compensate one with the other and give us a stable energy source?

Obviously, just combining randomly the solar and wind generation from one or different places will not give us any important conclusion, this is due to main facts:

The first one is that even if we combine sun and wind we will need to know when do we want them to produce at the same time for an increased production, when we are fine with a regular generation which can be produced by anyone of them or if there are times that they are not really needed so it can be whatever generation we obtain. All of this is determined by the energy demand we will want to cover; the total demand will of course be different for one city or the whole country and thus, the ideal combination of values will vary.

The other important point to take into account is the fact that the sun and wind conditions vary from place to place, therefore making some places generally better than others due to the fact of having better demand adapting sun and wind profiles.

These two points are the ones that in the following practical part will be studied and answered.

4. Practical application of the demand coverage qualitative study

In this first part of the practical part we will be doing a qualitative study of the relation between the solar and wind profiles from different locations of Spain and overseas, as well their combination and how good they fit a certain demand, in this case the one from the Iberian Peninsula.

To do this we will begin by gathering hourly data for solar and wind generation as well as the hourly energy demand, then we will be scaling it by their maximum values and create coefficients that we will call ponderation values. After that we will be using these values to do the qualitative study of the adaptation to the demand and draw some conclusions.

4.1. Selecting the data

The first thing we will have to take care of will be to know what kind of data we will be working with. As said before we will be studying the decorrelation of the sun and wind energy and how do they work together to cover their weak points. So the first step on to achieving this objective is to select the correct data for the study.

Obviously we will be studying the solar and wind power relation with the demand, so we will need some solar, wind and demand data sets, but the question is: How many sets, what kind of sets and how much precision do we need?

First of all a demand data set will be needed, so we can compare the values obtained by combining the solar and wind with something. For this project we will be using the energy demand in all the Iberian Peninsula between the 01/01/2007 up until the 31-12-2016; meaning 10 years of data will be studied. The demand data has been extracted from the *"Red Electrica Española"* webpage [14], downloaded as a .csv file and merged into an hourly demand. This data will be presented in one row for each hour of the data set; its units will be Megawatts (MW), therefore, because we are using a power coefficient by hours we can call it either power demand (MW) or energy demand (MWh)

As for the energy generation data, the first important thing to discuss is how many data sets we will need and of what kind of generation they will be. For this we have several options:

- Choosing only 1 energy source (solar or wind) or 2 or more of the same type only: This is not a good idea, basically because the whole objective of this thesis is to study the storage based on the combination of solar and wind power so they can compensate their unreliability.

- Choosing more than 1 sun data set and any number of wind sets: This is also a bad idea because the sun has basically the same shape in every place we can see; the value of the radiation may be different (and thus the total energy generation) but its behavior is almost identical in every place; we begin to have sun radiation at 6 a.m. and stop having it at 6 p.m. more or less; and for the rest of the day we have nothing. This means that we have that solar energies are simultaneous between themselves, which we don't want; therefore only one solar energy profile will be chosen.

- Choosing 1 sun and 1 wind data set from the same place: This is not a bad idea, because we would be able to see how a certain place can combine its sun and wind to cover the demand, but we limit ourselves in this way; we have to remember that we are connected to the electrical grid, and supposing that there are no energy losses, we can connect a solar power plant and a wind one from any place (including, of course the same place) and be able to take advantage of different wind profiles that can exist in different places.

- Choosing 1 sun and 1 wind data set from two different places (solar for one and wind for the other): This is a good option because we can combine the solar of 1 place with the wind of another, but what happens with the wind is that its profile is more irregular than the sun, so various different winds may be able to help into adapting to the demand better.

Therefore, as a final selection, we will choose to create a data set consisting of 1 sun and 2 winds, and for the sake of simplification and not having to increase exponentially the number of studied data sets we will couple one wind generation with the solar, so we will be studying two data sets at once, one with sun and wind and another with wind only.

In this projects the locations that will be studied will consist of the following Spanish regions: Barcelona, Madrid and Coruña; we will also be using two overseas regions as ideal values that we know to have great sun and/or winds, these regions correspond to the Sahara Desert (we will call it Sahara for future references) and the Scottish Highlands (we will call it Scotland for future references).

These data sets will be coded the following way: *Set_Name_1-Set_name_2*, e.g. "*Barcelona-Madrid"*, where the first set name corresponds to the data set that has the solar and wind values while the second set name has only the wind value. As to avoid forgetting important results complementary sets will always be studied (even though not always plotted if their information is redundant); meaning that when we study the case of Barcelona-Madrid, we will also study the case of Madrid-Barcelona.

Another notation point is that in future references and in the code that will be used for the calculations the wind from the first data set (the one containing sun and wind) will be also referenced as “wind 1”; while the second one (the one that has only wind) will be also referenced as “wind 2”.

This solar and wind data has been obtained through the Photovoltaic Geographical Information System (PVGIS) [15], by selecting the points of study (in this case from Spain and the other overseas locations).

One important thing to take into account when selecting the data we are required to select certain parameters; for this project all the data sets have been downloaded using the following parameters:

- Hourly data
- Solar radiation data base: PVGIS-SARAH
- Starting year: 2007
- Final year: 2016
- Mounting type: Vertical axis
- Optimize inclination
- FV power
- FV technology: Crystalline silicon (value by default)
- Peak power installed [kWp]: 1 (value by default)
- System losses [%]: 14 (value by default)

After choosing the selected place, the downloaded data contains the hourly values of the solar power generated by peak power installed (in Watts generated for every kW of installed power). We also obtain the hourly values of the wind speed at 10 m of the ground, and thus we need to elevate them to the average height where the actual wind generators will be, in this case it has been selected a height of 120 meters.

After downloading all the necessary data we will need to transform all the spread data into single data set that will contain the solar generation, both winds and the energy demand, which we will call, as we said before *Set_Name_1-Set_name_2*.

4.2. Preprocessing

For the preprocessing, as well as all the other practical sections of this project we will be creating some programs using Python 3; all of these programs will be available in the *Annex 1: Python code*.

We will begin with the structuring of all the data sets; for that we will open each downloaded set and start eliminating useless information; leaving us with only the date and time for each data row, the solar power (that will be converted to MW produced by MW installed by a simple conversion factor), the hourly power demand, the wind speed at 10 meters, which will eventually be discarded for the wind speed at 120 m; where our wind generator will be. To calculate this speed from the 10 m value we will use the vertical extrapolation of the wind speed [16] by using the log law shown in *Equation 4-1*:

$$U(120) = U_{10} * \frac{\log\left(\frac{z_{120}}{z_0}\right)}{\log\left(\frac{z_{10}}{z_0}\right)}; z_{120} = 120\ m; z_{10} = 10\ m$$

Equation 4-1: Wind speed logarithmic conversion formula

And:

u_{10}: Wind speed at 10 m above ground

z_0: Terrain roughness

For the terrain roughness, in this case, because we will be using a variety of data sets, we have to choose a standardized parameter so we will not have to keep modifying the code for every set we introduce. In this case we will be using a roughness coefficient constant for every place of 0.1 m. This will add a small amount of variability to the overall final result, but the difference obtained between coefficients is not that big, so we can assume a certain margin of error for simplification's sake. This can be one of the improvement points on future line of works, that is to make a speed converser that takes into account not only the terrain of every point, but also some other parameters like obstacles which may exist in the zone, for example.

After that we will need to transform these wind speeds at 120 m to wind power generated by power unit installed, the same way we have with the solar power.

To calculate the wind power from the speed we will be using the turbine power curve of a 1 MW turbine, this will be done so when we multiply that value with a coefficient, this coefficient will be at the same time the number of wind turbines, and because each wind turbine is of 1 MW it will also be toe installed power, therefore we will not be bound by the number of wind generators that will be needed in case we want to make a power plant; we will have the total installed power needed and the distribution will be free to be chosen as it seems fit.

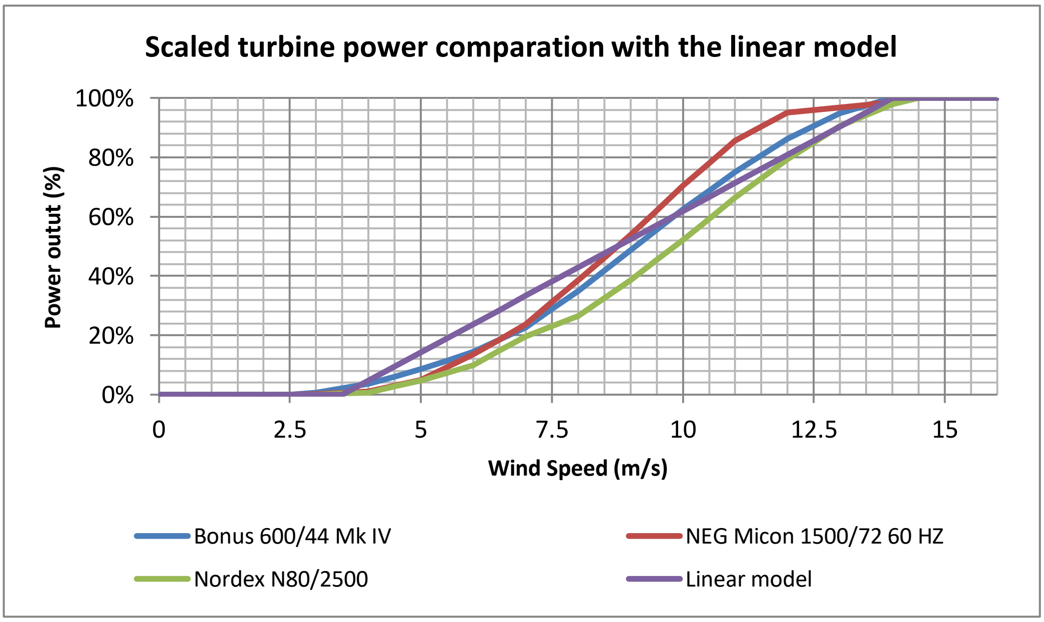
As for the approximation curve, because it is difficult to program a sigmoid, and using a data table to convert every wind speed value is very cumbersome and time consuming, we will be using a linear approximation of that said function which will be 0 at 3.5 and lower and 1 at 14 or higher up to a maximum of 25, which it will become 0 again due to the cut out speed. As we can see in the *Plot 4-3*; the linear estimation follows quite well the real curve.

Therefore, the mathematical formulation of the wind power as a function of the wind speed will be the one shown in *Equation 4-2*:

$$\begin{cases} P(v) = 0\,;\ v < 3.5 \\ P(v) = 0.0952 * v - 0.333\,;\ 3.5 \leq v \leq 14 \\ P(v) = 1\,;\ 14 < v \leq 25 \\ P(v) = 0\,;\ v > 25 \end{cases} ;\text{with } v \text{ in } m/s \text{ and } P\ in\ MW$$

Equation 4-2: Wind power calculation formula

As we can see in the following plot, if we compare the linear model created with the different scaled turbine power curves we saw in the theoretical part; we get a fairly good approximation, therefore, we will be using this equation from now on:



Plot 4-3: Scaled turbine power comparison with the created linear model

After having obtained all power generations from sun and wind, the next step is to get their relative values, which correspond on dividing the power value for sun and wind generations in every row by their maximum value of the column, so we have each generation in a scale from 0 to 1. We will also divide the energy demand in each row by the maximum value and multiply it by 3, therefore having the power demand scaled between 0 and 3. We will proceed to call these values ponderations, in this case sun ponderation, wind 1 ponderation, wind 2 ponderation and demand ponderation.

These ponderation values are very important for this section and we will take some time explaining them thoroughly.

4.2.1. Ponderation coefficients explanation

As we said before, we will be defining some coefficients we will call ponderations, these coefficients will help us do the qualitative analysis due to the fact that they are non dimensional (they are calculated by dividing a value by its maximum). This helps us because when trying to compare the energy generation with the demand we find ourselves looking at values with very different magnitude orders and therefore we need to scale all of them.

One important thing to mention is the fact that we are scaling the generations from 1 to 0, but the power demand is being scaled between 3 and 0. This is due to the fact that, because we are trying to cover the demand with all three generations by adding them, in the case that all three generations are on their maximum value we will have a total energy generation ponderation of 1 + 1 + 1 = 3; meaning that the global energy generation ponderation will range between 0 (when there is no generation at all from any source) to 3.

Therefore, we have to scale the demand between 0 and 3 because we are trying to calculate the hourly percentage of the demand that is covered; we consider that the maximum generation is the addition of the 3 different energy sources, initially all three with the same weight. Because we are comparing the energy demand with the total generation combination we need to have the total demand ponderation range between 0 and 3 if we have each generation range from 0 to 1. Another method which will yield the same results would be to scale the generations between 0 and 1/3 and the total demand between 0 and 1.

Finally as mentioned before we will calculate the hourly percentage of the demand covered by the generation; this will be done by just dividing the total general ponderation between the total demand ponderation obtaining values under 1, if the demand is not completely covered, 1 if it is and over 1 if there is an energy excess.

It's important to note that not only the demand will be covered when all the generations are maximum and the demand is also maximum; but for example, if we had a demand half as big as the maximum value (with a ponderation coefficient of 1.5) and we have max solar power, none of wind 1 and half of wind 2 (giving us ponderation values of 1, 0, 0.5 respectively) we would still have all the demand covered (because $\frac{1+0+0.5}{1.5} = 1$). The same happens for the excess and deficit coverage.

4.2.2. Final data set format

Once we have finished calculating all the needed parameters we can create the data set that we will use for all the future practical studies; this data set contains both sun, wind 1 and wind 2 generations, the energy demand the power generated values as well as their respective ponderations. This full data set contains the following information:

- Date in the format: "yyyy-mm-dd hh" plus the set name
  - Column name: "*Set_Name_1/Set_name_2:time*"
- Solar power in MW generated per MW installed
  - Column name: "Sun_power"
- Solar ponderation
  - Column name: "Sun_pond"
- Wind speed at 120 m from the first set (solar + wind) in m/s
  - Column name: "Wind1_speed"
- Wind power from the first set in MW generated per MW installed
  - Column name: "Wind1_power"
- Wind 1 ponderation
  - Column name: "Wind_pond"
- Wind speed at 120 m from the second set (wind only) in m/s
  - Column name: "Wind2_speed"
- Wind power from the second set in MW generated per MW installed
  - Column name: "Wind2_power"
- Wind 2 ponderation
  - Column name: "Wind2_pond"
- Power demand in MW
  - Column name: "Demand"
- Demand ponderation
  - Column name: "Demand_pond"
- Total generation ponderation
  - Column name: "Gen_pond"
- Demand coverage ponderation
  - Column name: "Dem_cov"

This data set will be the same we will use in the future storage study we will do; from this point forward, when we reference a set, we will be talking about a set with the format just showed, with one solar and two winds generations.

With the data set creation complete we are ready to begin the qualitative study, or as we will also call it, the ponderation study and compare the data sets against each other.

4.3. Hourly ponderation study

As said before, we will be using the ponderation values calculated before to do a qualitative study on how well a given demand is covered by the combination of sun and different wind generations. To be able to see the relation and the demand coverage a different series of plots will be created, as well a comparative coefficient that will help us to get a first approach of the better sets. The plots that we will be using in this first part are the following:

- The first plot will consist of the additive sum of the solar wind 1 and wind 2 (in three different lines colored yellow, blue and cyan respectively) of the generation ponderations defined before and the demand ponderation, so we can see how much each generation covers each hour. Because we are working with such a big amount of data, it is physically impossible to fit all the 10 years hour by hour in a plot, therefore we will be using the mean value of each ponderation at the same hour for all the ten years, meaning that we will have a plot with the X axis divided in 24 sections (from 0h to 23h), and thus, obtaining then the overall mean value of the ponderation generations and demand.

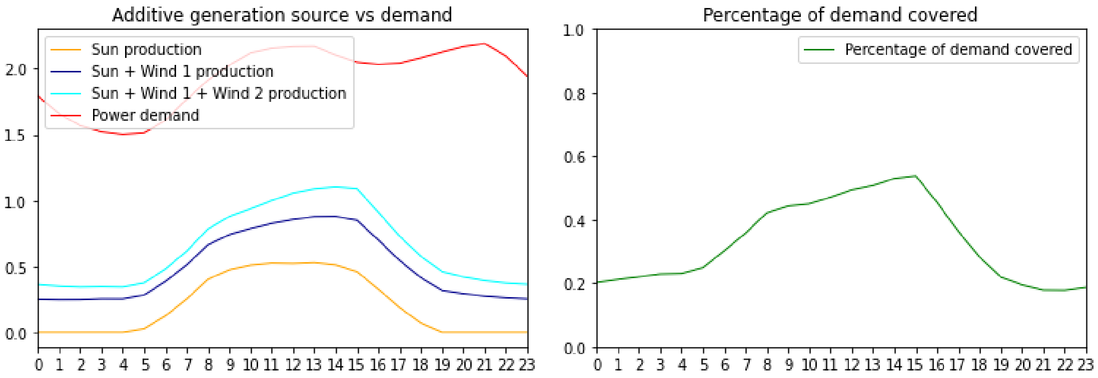
  It is important to note that because each generation line is the sum of itself and the previous one, the cyan line will also represent the total generation ponderation of that set.

- The second plot will be the total demand coverage percentage, obtained from dividing the total generation ponderation (the cyan line) between the demand ponderation (the red line), which will tell us how well the demand adaptation is. In this case we will also be condensing all the 10 years of data into hourly mean values (from 0h to 23h).

In this section the data sets we will be using and plotting are the combination of some of the following data sets: Barcelona, Madrid, Coruña, Sahara and Scotland. To be more precise the sets that we will be plotting results will be: Barcelona-Madrid, Madrid-Barcelona, Barcelona-Coruña, Coruña-Barcelona, Sahara-Scotland and Scotland-Sahara. After that, when discussing the results some more combinations will be presented to widen the conclusions, these will consist of other combinations between some of the sets mentioned previously.

4.3.1. Barcelona-Madrid data set

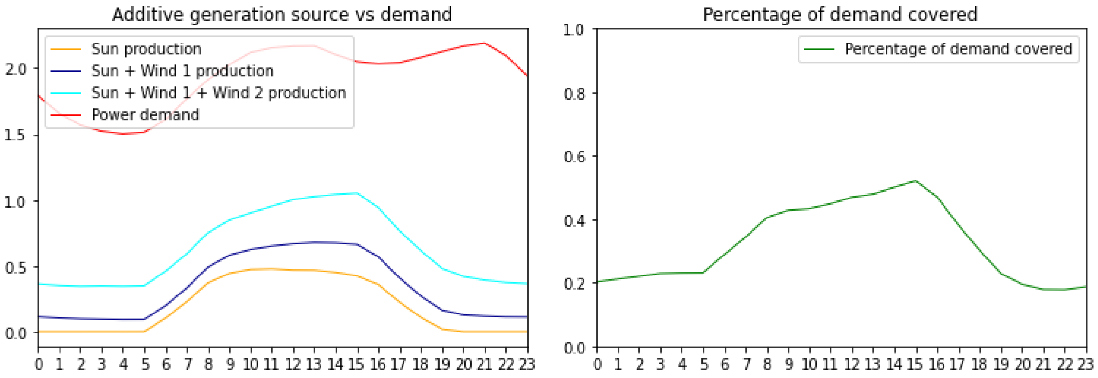
The following two plots correspond to the additive generation vs demand ponderations and the mean hourly demand coverage.



Plot 4-4 and 4-5: Additive generation source vs demand and demand covered percentage for the data set

4.3.2. Madrid-Barcelona data set

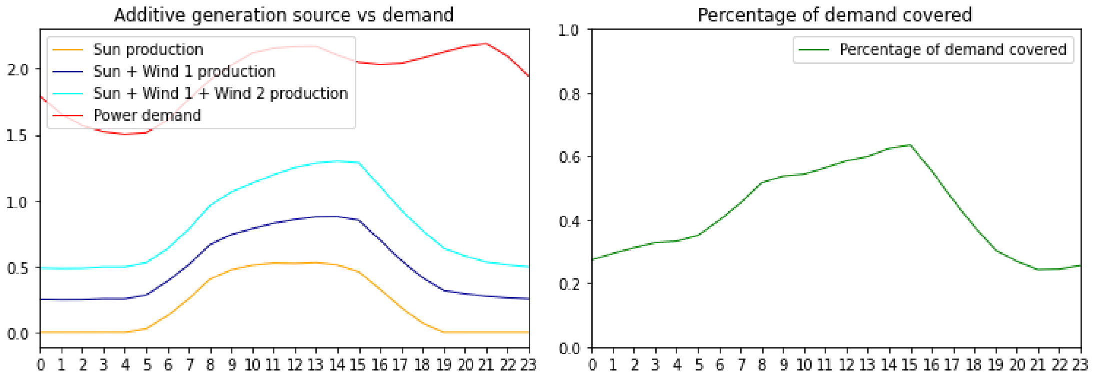
The following two plots correspond to the additive generation vs demand ponderations and the mean hourly demand coverage.



Plot 4-6 and 4-7: Additive generation source vs demand and demand covered percentage for the data set

4.3.3. Barcelona-Coruña data set

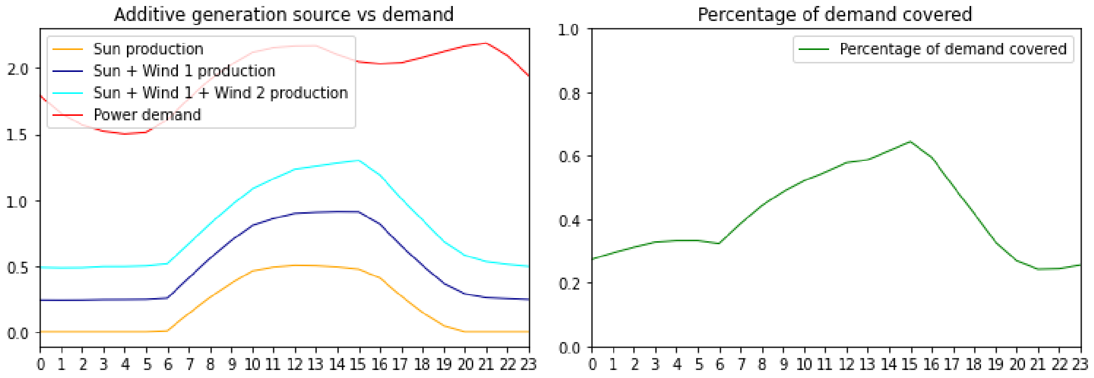
The following two plots correspond to the additive generation vs demand ponderations and the mean hourly demand coverage.



Plot 4-8 and 4-9: Additive generation source vs demand and demand covered percentage for the data set

4.3.4. Coruña-Barcelona data set

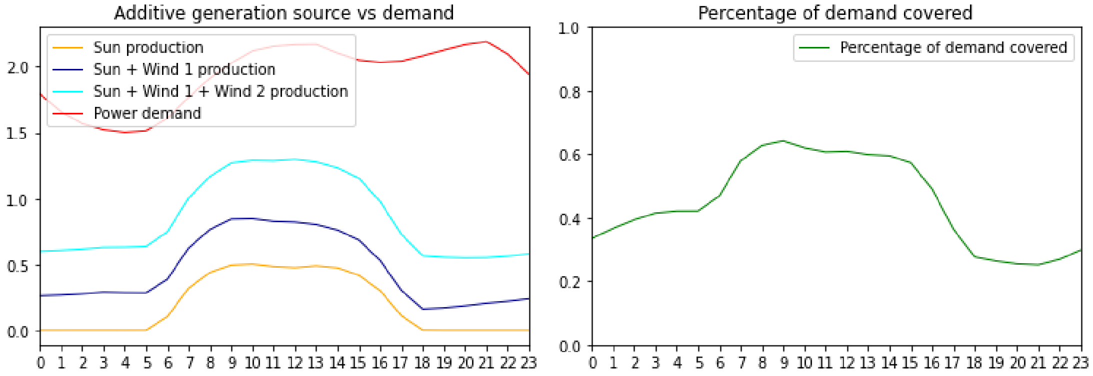
The following two plots correspond to the additive generation vs demand ponderations and the mean hourly demand coverage.



Plot 4-10 and 4-11: Additive generation source vs demand and demand covered percentage for the data set

4.3.5. Sahara-Scotland data set

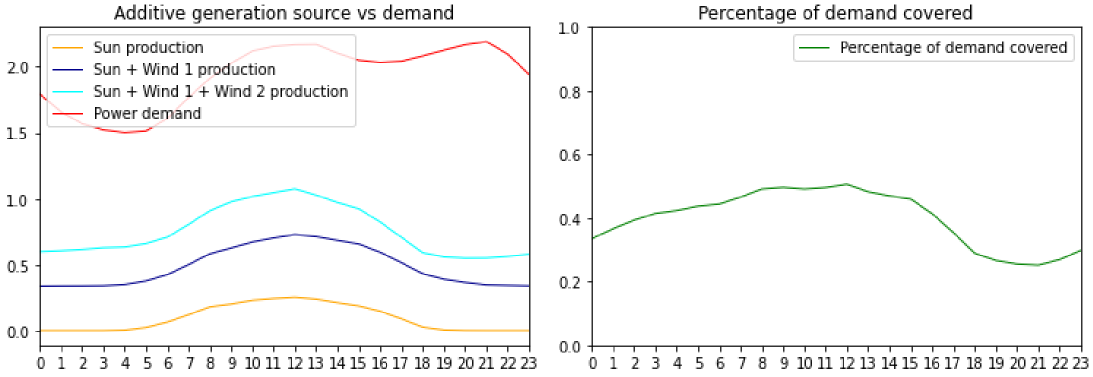
The following two plots correspond to the additive generation vs demand ponderations and the mean hourly demand coverage.



Plot 4-12 and 4-13: Additive generation source vs demand and demand covered percentage for the data set

4.3.6. Scotland-Sahara data set

The following two plots correspond to the additive generation vs demand ponderations and the mean hourly demand coverage.



Plot 4-14 and 4-15: Additive generation source vs demand and demand covered percentage for the data set

4.3.7 Results analysis

When looking at the obtained plots we can see the different behaviors that different data sets provide.

The first thing that we can see is that in all the cases, the sun production is high at the moments the demand is also high; which is a logical thing because if we take a look at reality, the maximum demand hours are when everybody is at home and not sleeping; namely the morning and afternoon. This is why we can see that the demand coverage is always higher during sun hours due to the extra generation it provides.

Another thing that comes right in sight is the differences between the solar generations in various places; we can clearly see the difference between the sun from Scotland and Sahara (*Plot 4-12* and *Plot 4-14)*; one being almost nil while the other is very high. Another comparison can be done with the sun from Barcelona and Coruña; we can see that in these two sets (Barcelona-Coruña and Coruña-Barcelona) both wind coefficients are the same (*Plot 4-8* and *Plot 4-10)*, so the only difference can be seen in the solar generation. While in the additive generations maybe we can not see the difference, in the percentage of demand covered we can clearly appreciate the difference between them; the hourly coverage in morning hours is better in Barcelona than in Coruña. As for Madrid and Barcelona we can see that both plots are almost identical (*Plot 4-4* and *Plot 4-6)*, this is due to the fact that both Madrid's sun and Barcelona's is basically the same.

As for the wind we can see that most of them have important values except Madrid, whose values are fairly small; this leads to a lower demand covered percentage plot that the rest of data sets; about this case more will be explained later.

Still, all these results are fairly qualitative, so in order to obtain a somewhat objective value that allows us to know which combination of places is better at covering the demand we will do the mean value of all hourly demand coverage percentages; this will give us the overall coverage percentage, which will use to compare sets. The values of those coefficients from the plotted sets before, and some more also studied will be shown in the following *Table 4-16.*

Set Name	Overall coverage (%)	Set Name	Overall coverage (%)
Coruña-Scotland	46.44	Madrid-Coruña	33.16
Barcelona-Scotland	46.43	Coruña-Madrid	32.97
Sahara-Scotland	44.73	Barcelona-Madrid	32.96
Barcelona-Coruña	41.88	Madrid-Barcelona	32.33
Coruña-Barcelona	41.06	Madrid-Sahara	31.29
Scotland-Barcelona	40.89	Sahara-Madrid	31.12
Scotland-Sahara	39.85		

Table 4-16: mean hourly demand coverage percentages for the studied sets

We see that the coefficients with a bigger overall coverage are those that have higher and more constant generations either of sun or wind (like Sahara, Coruña, Barcelona and Scotland). We can also see that the wind is a very influential value because we are calculating the total demand coverage by using two winds and one sun, all with the same weight, therefore, if the set has two strong winds values we will have a higher coverage, even if we have a weak sun (like Coruña-Scotland).

The same way applies to sets that have one or more weak coefficients, like Madrid, in this case because all coefficients weight the same we get a smaller coverage due to a third or more of the total generation value being very small, and thus we obtain a smaller generation ponderation.

We will see in the future, that this relatively high coverage can be detrimental if only the winds are powerful, because our objective is, as we said in the theoretical part of this project, is to reduce the unreliability of the wind and sun by combining them with each other, and thus if we were only to look at a 100% wind generation, the whole premise of this project will be invalid, because we will just be doing a wind power study. Therefore one important hypothesis we can make, and will be studied more thoroughly in the following sections of this project is that:

Even though a good coverage can be achieved through powerful winds only, if we have to combine them with sun generation and that generation is weak, we will find ourselves with worse results than slightly smaller but better distributed solar and wind values.

Obviously there are situations where both sun and wind generations are high (maybe not as big as other winds, but still relatively big) so they have also higher coverage values; in this case we are talking about the Barcelona-Scotland, Barcelona-Coruña and Sahara-Scotland.

Finally we have those smaller values, those which mainly involve Madrid. This is due to the fact that, as we have seen in the plots, the Madrid wind coefficient is very small and therefore the overall coverage is smaller. This can be seen first as a disadvantage, because even if a location lacks high values for winds or sun to cover the demand, if we use it as the support set of another one with higher values, even though we will be needing more installed power to be the same as others that will require less, we can end up obtaining a better demand covering overall.

This will also be especially relevant when we do the future storage study and present installed power values rather than this qualitative approach. Therefore another important hypothesis we can make, and will be studied more thoroughly in the following sections of this project is that:

Even if we have a small overall coverage value, if we combine the different locations cleverly, we can still obtain good power demand coverage.

4.4. Overall conclusions of the qualitative study

After doing the qualitative study of the demand coverage we now have some better understanding of the influence that each factor has in the demand covering; but because we have treated every generation with the same weight we obtained some results that may not be intuitive at first; even though we have explained, at least qualitatively some reasons for that behavior; obtaining the following hypothesis:

Even though a good coverage can be achieved through powerful winds only, if we have to combine them with sun generation and that generation is weak, we will find ourselves with worse results than slightly smaller but better distributed solar and wind values.

Even if we have a small overall coverage value, if we combine the different locations cleverly, we can still obtain good power demand coverage.

These two new hypotheses, which we have defined in this section will be some of the questions that we will ask ourselves in the following section of this project; the quantitative energy storage study. But before proceeding to that section we will do a little review on what defines a storage system and the different implementation methods that exist, along with some important factors to take into account.

5. Energy storage

In this section we will be doing an introduction on what will be the second part of this project; which will focus on the energy storage of the solar and wind mixed production. As one may know, the storage of energy consists on the capture of energy to use at a later time [17]; when the energy production is lower and the demand still has to be covered.

There are plenty of different energy storage types; some of them are more short term oriented, while others may last for a longer time; in the following point we will discuss some of the most common ones.

5.1. Types of energy storage

There are several types of energy storage; we will proceed to discuss the most common ones.

5.1.1. Fossil fuel storage

Fossil fuels like coal, oil or natural gas are by themselves energy storages; they contain energy created in ancient times by the death and decomposition of organisms. This is one of the strong points of these types of fuels, due to being able to be stored for indefinitely amounts of time, and used when necessary without having to worry about energy losses in the waiting period.

5.1.2. Mechanical storage

The storage of energy via mechanical means consists on the transformation of excess energy into potential or kinetic energy for the future use.

The most well known method of this type of storage is the pumped hydro, which consists on the storage of energy in the form of potential gravitational energy pumped from a lower reservoir to a higher one. This pumping is done when the cost of the electricity in the market is low, in other words, when the energy demand is low. After that, when more energy is needed to supply the grid, the system works as a normal hydro electrical power station.

This is the most widespread storage system in the whole world, representing up to the 95% of the total installed storage capacity all over the world, due to the fact of having a very high efficiency and long term storage capabilities [18].

Other examples of mechanical storage used more in local scenarios and more short term periods is the use of compressed air or flywheels to store kinetic energy for later use.

5.1.3. Electrochemical storage

Electrochemical storage systems, better known as batteries, are short term storage systems that contains different cells with different chemicals that, when combined together produce a chemical reaction that generates electricity. Nowadays, most of these batteries are rechargeable, meaning that if given electrical energy they can undo the chemical reaction of the components and leave them for a future use. These batteries can be made of a lot of different materials like lead, nickel, cadmium or more recently lithium.

5.3.4. Other types of storage

A part of the mentioned ones there are several other types of storage; some of them are:

- Chemical storage: Like hydrogen, methane or food.
- Electrical storage: Like capacitors or superconductor magnets
- Thermal: Like molten salts or heating accumulation towers

5.2. Energy storage selection

In the case of this project, because sun and wind are unreliable sources of energy and they can't be physically stored for future use, the energy storage for the electricity generated becomes even more necessary.

As we have seen there are plenty of different ways to store energy, also, in this next section of the project we will try to calculate the storage needed for the combined generation of sun and wind; therefore one question that may arise would be what kind of storage we will use for this task.

However, as we mentioned in the project's objectives, our main focus is to study how sun and wind are able to work together to compensate their inherent unreliability and cover the demand; therefore, the only thing we need from that storage is its value so we can compare it with other different locations, for this reason we will not be choosing any specific storage type.

If we were to take into account the total cost of the installed power and storage system, it would be mandatory to choose a valid storage system depending on the obtained results, because if for the obtained storage we needed to store energy for long periods of time, systems like batteries will not work; the opposite is also true, if we were to only have small short term requirements, storages like hydro pumping would be an excessive option, due to its high cost and capacity. This uncertainty and the lack of relevance in the project's goal are the reasons that we will not bother choosing a concrete storage type.

With that said, in the next section we will be defining a storage system and selecting coefficients for the installed power we will need to cover the required demand, then we will try to minimize its overall capacity to find out which is the best location for said installation.

6. Practical application of the storage study

6.1. Practical application introduction

In this second part of the project, as mentioned before in the theoretical explanation; a storage study will be carried out to see how the decorrelation of the renewable generation affects to the overall storage capacity; either with its maximum value and its coefficients or with the excess/deficit of energy it creates when covering the hourly demand.

This will also help us confirm the two hypotheses presented in the ponderation study, due to the fact that in this case we will not be considering every generation of the same weight we will be able to see their true coefficients and the best way they will adapt to the demand.

In this second practical section, in order to do all the necessary calculations and plots we will be using Python 3; all of the code used in this art will be available in the *Annex 1: Python code*.

In this second part, as opposed to the ponderation study, we will be using real values with power or energy units, in this case megawatts (MW) and megawatts per hour (MW/h). This means that we will be looking at the energy demand of the whole country of Spain (which is the one we got the data from) and facing it with an energy generation from two chosen locations.

This can be seen as a problem magnitude wise; because it ends up comparing values of the thousands of megawatts with a small local generation, and thus when the time comes to assign the final installed capacity for each generation source (solar and both winds) we can end up seeing values of high dimensions (up to 10^{4-5} MW of installed power and even more). These values however, are actually logical, due to the fact that we are considering the fact that we are covering the whole country's energy demand from two single points, and thus, we will need an abnormally large installed capacity (which in reality doesn't need to physically fit in the selected locations).

As mentioned before, when searching for the optimal storage, we will be multiplying the base generations of MW generated per MW installed by certain coefficients so we get the installed power value and adapt it to the hourly demand. After that we will choose the appropriate storage capacity for each place.

To calculate the final values of the installed capacity, an optimization program has been used, more specifically a minimization one, the process used in the minimization has been as follows:

- We begin by separating all the data by weeks (in groups of 7*24 = 168 rows which represent the hours of that specific week).

  A weekly separation has been chosen due to the fact that if we were to create a daily separation, (which the implemented program can work with), we will find ourselves with an extremely long iteration times, because all the calculations would end up multiplying seven times, which is heavily memory and time consuming.

  We didn't choose a monthly separation either due to the fact that at that point a lot of information would be lost, also climatological patterns are more common to change from month to month, than from week to week, and therefore we chose the weekly more balanced approach which combines precision and speed.

- For each week and some initial parameters of installed power we determine the hourly energy generation and subtract it from the hourly demand, thus obtaining the hourly energy deficit (positive when there's energy excess and negative when there is a deficit).

- After that we calculate the accumulative sum of the deficit by adding each deficit value to the previous one, which makes us able to know how the total energy excess or necessity varies according to previous values.

- Then, we choose the maximum absolute value as the maximum storage which will be needed that week (if the value comes from a positive one means that the maximum is obtained through an excess and if it's negative the maximum is obtained from a deficit).

- Finally we run the minimization program, where we try to minimize that daily storage by iterating with the installed power coefficients with the restriction that every week total generation (the sum of it) must be equal to the weekly total energy demand.

  Some considerations to take into account when doing the minimization are the values of the boundaries for the estimated coefficients; the first and foremost is that they can not be under zero, this is due to the fact that even if we get a better approximation (due to the fact that sometimes combining negative and positive coefficients can make better regressions), a negative energy generation is not only not viable but also illogical in engineering terms.

Once these values are obtained, however, there are more things that needed to do before defining a definitive storage and installed powers. Because the values obtained before are different for each week and the installed power can't be changed weekly, therefore, we need to define some final values to install that they will remain constant throughout every week, and thus every hour. Also, none of the maximum storages minimized before will be the ones chosen as the final one due to the fact that each and every one of them has been chosen using different coefficients. In the next step we choose the installed power coefficients which will define the final value of the storage.

We can't just choose the value corresponding to the minimum storage needed on a certain week due to the fact that every week the optimization is done using the values of the sun and wind for that day, and every day has different values, so it's very possible that the parameters for a week might have, for example, a good storage value obtained by using only the second wind generation and disregarding the sun and the other wind values, due to the fact that that week the first wind adapted perfectly to the demand. If we were to use these values for every week and therefore for every hour, we will very likely find ourselves in a situation where most of the days that depend on other parameters are on a great deficit.

Therefore, we will be calculating the average value of all the different coefficients obtained every period of time so in this case, even though we may have lower coefficients, and thus, a lower generation than we would have if we used the maximum value of the coefficients we will avoid having extreme values which can end up giving us excessively large storage capacities. This means that we will be calculating the mean of the solar and each wind coefficients and choose those three values as the final installed coefficients for each respective generation.

This has been done because there are a lot of different parameters that can affect the final values of the coefficients; but the most important one are the boundary conditions. This is due to the fact that the minimization algorithm is defined between two boundaries, the upper and the lower. In case of the lower boundary, it just has to be zero so we don't have any negative value. But when we are looking at the upper one, the value we choose there will determine basically everything, meaning that depending on the value given the coefficients can change by orders of magnitude, giving us very different generation coefficients and, because is strongly related, end up affecting the storage values. That is the reason that very high upper threshold has been chosen, so we have little to no coefficients affected by the boundary conditions and, by doing the average, we make double sure that we eliminate the extreme values themselves.

After obtaining the final coefficients of installed power, the next step is to define a maximum storage capacity. This will be done in a similar way the minimization program was used before; but in this case, as we already have the coefficients we will be calculating the maximum weekly storage needed, which then will be plotted in a histogram; then from all the storage values of every day, one value will be chosen as the maximum storage capacity depending on the percentage of storage we are willing to renounce, so we can make the storage more feasible.

This is due to the fact that because we are using high values, the energy generation will be on the large side of the spectrum, and thus there will be a lot of excess for some days, meaning we will need a very large storage to accommodate all that energy; so we sacrifice some of this high storage values in front of a more realistic value and the fact that the energy excess can be sold off and the possible deficit can be countered by other punctual energy sources (like combined cycle, coal or from a more ecological point of view, hydro or biomass).

The choosing of the storage with this conservative and tending to the maximum criteria is done due to the fact that the reach of the study doesn't take into account the cost of the energy, only the decorrelation from the coefficients and how they translate into the storage value, so we will not be focusing on the excess of energy generated (which will be over the capacity of the storage system, and thus will have to be dealt with), but rather the storage itself and how it will vary depending on the chosen coefficients.

Even though we are using a conservative criterion which tries to equalize the total weekly generation and weekly demand, there may be some periods of time that the energy generation will not suffice to adapt to the demand and thus, there could be an energy deficit, so it's important to know how long will that deficit last, because long periods of time without having any kind of energy supply can be catastrophic. So as the last part of the storage analysis excess and deficit energy streaks will be calculated so we can see it the time frames we are dealing with when we are over the maximum capacity or not even reaching to cover the hourly demand.

As a final point of this section it is worth to mention that the values obtained for the generation coefficients are not really optimal, due to the fact that we are using a hand picked value for the thresholds and the method for selecting the coefficients for the installed power can be seen as somewhat heuristic.

Despite all that, as we commented before, the main objective is to be able to see how the solar and wind profiles, and therefore their generation influence the storage capacity, and thus, because we are using the same boundary conditions for all the different places, and the values that define the decorrelation (which are sun and the two wind generations) are not affected by these conditions, we can fully analyze the storage system and draw conclusions without fear of losing important information.

To sum it all up, in the next section we will be calculating the storage data and relevant parameters of the fictitious storing we have created to check how it varies due to the solar and wind decorrelation. To do so we will be dividing the results in 3 main parts:

The first part will be the calculation of the storage value and the generation coefficients for each solar and wind power. Next we will study the value level of that said storage and how many hours we are having an excess of energy, a deficit or if we are between thresholds of the storage (meaning its still not fully charged or discharged), this will help us see how do the chosen coefficients affect the energy distribution and the number of hours in each of the three sections (excess, deficit or threshold). Finally, the last part will consist on the study of the aforementioned excess and deficit hours and how are they distributed along all the data.

6.2. Implementation of the practical approach

The first step on this second part is to select the study elements and to define the parameters so we will not have undesired uncertainty when obtaining the results. This is, in this case, of special importance, because as mentioned before the storage calculation in not optimal, due to the fact that it uses some heuristic methods.

Another point to take into account is the fact that this storage value could be optimized through multiples iterations; first calculating the coefficients, then the storage and once we have the storage use it to recalculate new coefficients and so on. This iteration method hasn't been implemented due to two main reasons. The first one is due to memory limitations and time consumption, due to the fact that the storage calculation has been done week by week, so we have obtained around 520 storage results and coefficients (one for every week for ten years), then to iterate we would be needing to choose a new storage value for each of those coefficients and selecting the optimal one, which will be incrementing the number of calculations exponentially. The other reason is what we mentioned before; the objective of this second part is not to find the ideal storage value and coefficients, but to see how it gets determined by cause of the sun and wind disparity, so even if we don't get the perfect values we will still be able to see the results we wanted.

In this section the data sets we will be using and studying are the same ones that we used in the decorrelation part: Barcelona, Madrid, Coruña, Sahara and Scotland, and as we did before we will be comparing the sets in the following order: Barcelona-Madrid, Madrid-Barcelona, Barcelona-Coruña, Coruña-Barcelona, Sahara-Scotland and Scotland-Sahara. After that, when discussing the results some more combinations will be presented to widen the results, these will consist of other combinations between some of the sets mentioned previously.

6.3. Storage values and coefficients calculation

First of all, we will proceed to define the initial hypothesis and threshold conditions used in this section which will be the same for every studied case to avoid undesired variability. These are the following:

- For the minimization part we will be using boundary conditions for the generation coefficients values (the same for solar and wind generation); the lower boundary will be the value 0, due to the fact, already mentioned before, that we can not have a negative energy generation. The upper threshold will be the value of 1 000 000, this is due to the fact that we don't want a limitation of the coefficients, or at least to have the least influence on the obtained results, their initial values will be 10 for all of them.

- The minimization function will be the following one:

$$\min(\max(\, cumsum(deficit)))$$

Where:

$$deficit_i = gen_{solar,i} * c_{solar} + gen_{wind1,i} * c_{wind1} + gen_{wind2,i} * c_{wind_2} - demand_i$$

$$\forall i \in week$$

$cumsum \rightarrow$ Accumulative sum of the data values

$gen_{solar,i} \rightarrow$ Solar base value for hour i in the corresponding week subset

$gen_{wind1,i} \rightarrow$ Wind 1 base value for hour i in the corresponding week subset

$gen_{wind2,i} \rightarrow$ Wind 2 base value for hour i in the corresponding week subset

$c_j \rightarrow$ Coefficient corresponding to the installed power for each generation type

Following the restriction:

$$\sum_{week} gen_{solar} * c_{solar} + +gen_{wind1,i} * c_{wind1} + gen_{wind2,i} * c_{wind_2} = \sum_{week} demand$$

- The data grouping for the minimization process will be done in week by week subsets, which meant that each subset will have 24 * 7 = 168 rows (each one representing 1 hour) and a total of 52 * 10 = 520 subsets (52 weeks every year).

- The percentage we are willing to renounce when choosing the storage is 25%, meaning that after calculating the coefficients and using them to obtain all storage values; we will be using the maximum value from the 75% lower percentile of all the values.

After having explained the procedure in this section we will proceed to show the data corresponding to the sets mentioned before, after that we will proceed to obtain conclusions from the data.

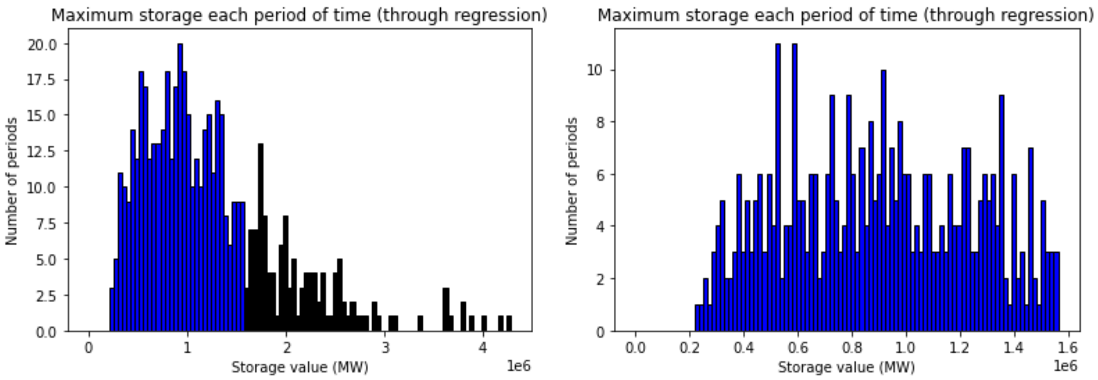
6.3.1. Barcelona-Madrid data set

After running the minimization program we obtain that the values for the installed power values:

Solar coefficient (MW installed)	**Wind 1 coefficient (MW installed)**	**Wind 2 coefficient (MW installed)**	**Storage value (MWh)**
78 530	**24 721**	**36 548**	**1 561 210**

Table 6-1: Energy generation coefficients and storage value chosen

In the next two plots we can see the storage values calculated with the obtained coefficients for each period (in blue), as well the storage values we have renounced (colored in black). The rightmost value of the blue bars represents the storage value chosen and shown in the table before. The first plot corresponds to all the storage values while the second one corresponds to the expanded values of the blue bars of the first plot.



Plot 6-2 and 6-3: Maximum storage each period of time with and without the renounced hours

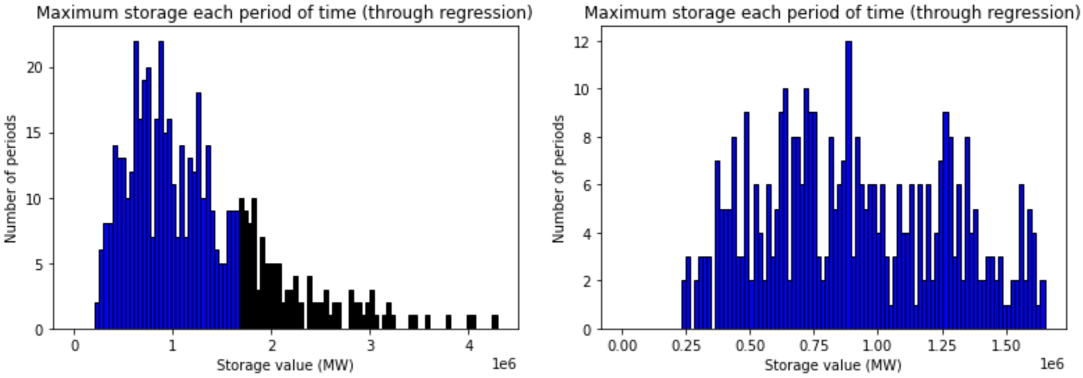
As to not be overly repetitive, and because the plots are giving the same information for different sets, the explanation done before will not be repeated for the further results.

6.3.2. Madrid-Barcelona data set

For this set we plot the same information mentioned before, the data table and the storage plots.

Solar coefficient (MW installed)	Wind 1 coefficient (MW installed)	Wind 2 coefficient (MW installed)	Storage value (MWh)
77 911	**36 960**	**24 876**	**1 651 314**

Table 6-4: Energy generation coefficients and storage value chosen



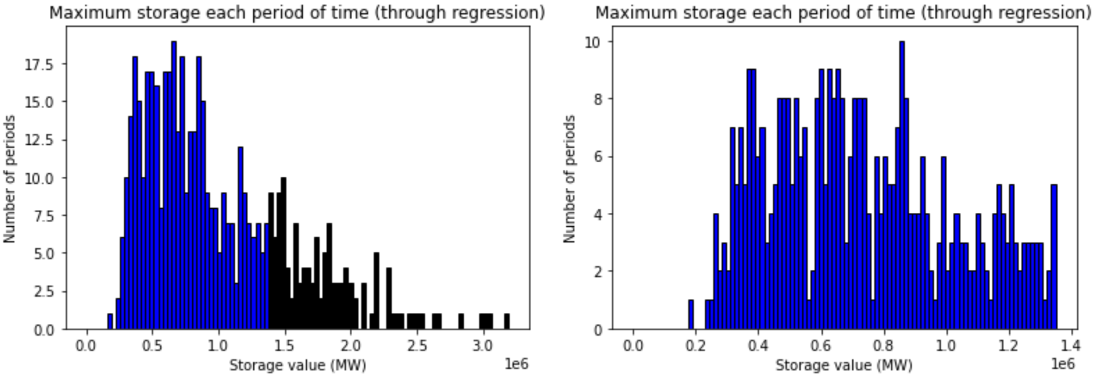
Plot 6-5 and 6-6: Maximum storage each period of time with and without the renounced hours

6.3.3. Barcelona-Coruña data set

For this set we plot the same information mentioned before, the data table and the storage plots.

Solar coefficient (MW installed)	Wind 1 coefficient (MW installed)	Wind 2 coefficient (MW installed)	Storage value (MWh)
66 155	**25 450**	**20 071**	**1 348 914**

Table 6-7: Energy generation coefficients and storage value chosen



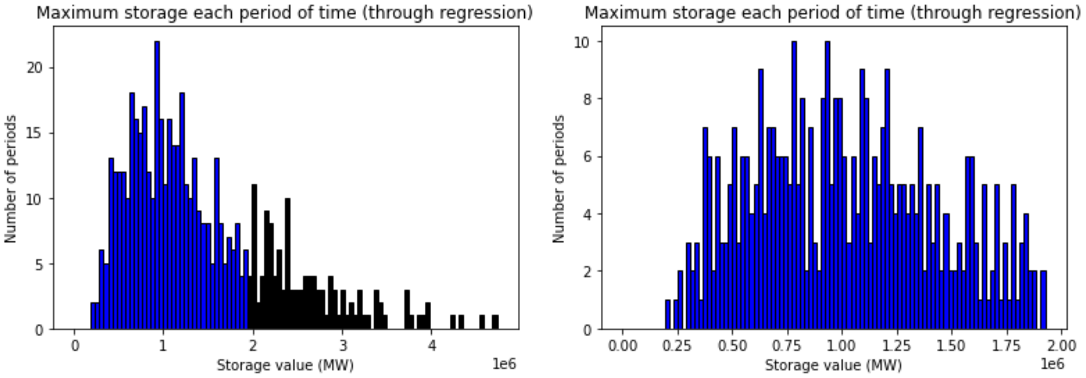
Plot 6-8 and 6-9: Maximum storage each period of time with and without the renounced hours

6.3.4. Coruña-Barcelona data set

For this set we plot the same information mentioned before, the data table and the storage plots.

Solar coefficient (MW installed)	Wind 1 coefficient (MW installed)	Wind 2 coefficient (MW installed)	Storage value (MWh)
77 102	**32 119**	**33 767**	**1 924 883**

Table 6-10: Energy generation coefficients and storage value chosen


Maximum storage each period of time (through regression)
Maximum storage each period of time (through regression)

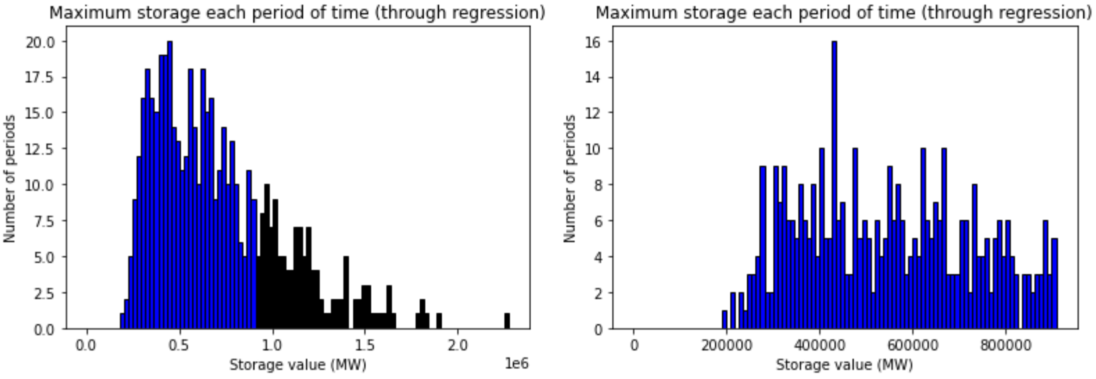
Plot 6-11 and 6-12: Maximum storage each period of time with and without the renounced hours

6.3.5. Sahara-Scotland data set

For this set we plot the same information mentioned before, the data table and the storage plots.

Solar coefficient (MW installed)	Wind 1 coefficient (MW installed)	Wind 2 coefficient (MW installed)	Storage value (MWh)
72 360	**16 702**	**10 784**	**952 203**

Table 6-13: Energy generation coefficients and storage value chosen


Maximum storage each period of time (through regression)
Maximum storage each period of time (through regression)

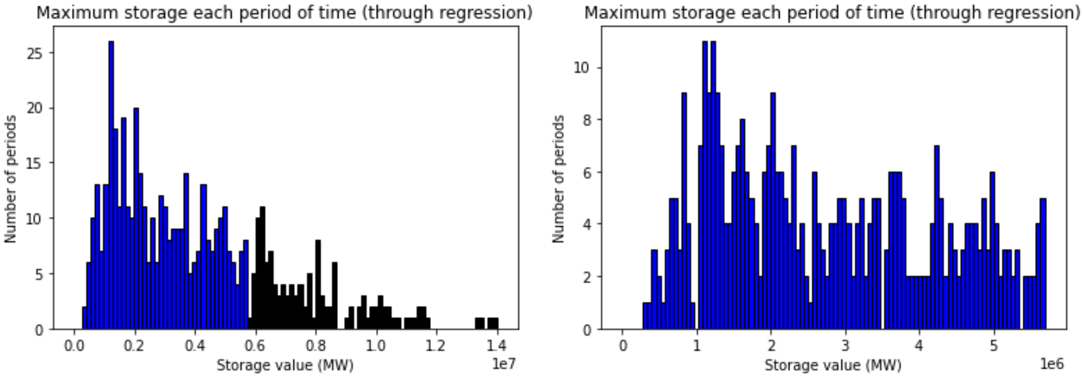
Plot 6-14 and 6-15: Maximum storage each period of time with and without the renounced hours

6.3.6. Scotland-Sahara data set

For this set we plot the same information mentioned before, the data table and the storage plots.

Solar coefficient (MW installed)	Wind 1 coefficient (MW installed)	Wind 2 coefficient (MW installed)	Storage value (MWh)
244 236	**31 894**	**45 820**	**4 942 264**

Table 6-16: Energy generation coefficients and storage value chosen



Plot 6-17 and 6-18: Maximum storage each period of time with and without the renounced hours

6.3.7. Results analysis

After having witnessed the results and plots we can proceed to note some important points. To help expand on these conclusions other data sets have been studied but its graphs haven't been plotted to avoid an excessive number of similar plots, even so, some of the relevant values of these extra data sets will be tabulated along with the previous coefficients for easy access:

Set (L = Local // O = Overseas)	Solar coefficient (MW installed)	Wind 1 coefficient (MW installed)	Wind 2 coefficient (MW installed)	Storage value (MWh)
Barcelona-Madrid (L)	78 530	24 721	36 548	1 561 210
Madrid-Barcelona (L)	77 911	36 960	24 876	1 651 314
Barcelona-Coruña (L)	66 155	25 450	20 071	1 348 914
Coruña-Barcelona (L)	77 102	32 119	33 767	1 924 883
Sahara-Scotland (O)	72 360	16 702	10 784	952 203
Scotland-Sahara (O)	244 236	31 894	45 820	4 942 264
Madrid-Coruña (L)	77 497	35 674	22 026	1 573 943
Coruña-Madrid (L)	97 583	36 339	52 558	2 462 086
Madrid-Sahara (O)	65 708	49 945	34 858	1 668 789
Sahara-Madrid (O)	76 445	17 417	27 587	973 221
Barcelona-Scotland (O)	57 770	29 681	17 432	1 197 648
Scotland-Barcelona (O)	276 183	29 140	39 243	5 264 722
Coruña-Scotland (O)	69 713	38 146	24 073	1 977 123

Table 6-19: Solar and wind coefficients and storage value from all the studied data sets (all units in MW)

After witnessing the results and different plots, we can begin to note some important points:

Regarding the storage value chosen plots we can see that we have taken out the tails of the graph (*Plots 6-2, 6-5, 6-8, 6-11, 6-14 and 6-17)*in order to get a smaller and more feasible storage value. This has been done to eliminate the outliers that we can see, exist in every plot; it is true that by doing so we are losing some values that may not be that big, but as we said before is a sacrifice we are willing to make. Also we know that in reality we could compensate that loss with other energy sources.

Diving into the coefficients, we can see that the sun value is larger that the wind values in all the cases; this is due to the fact that the sun has a very different profile than the winds, so it gets a bigger coefficient because it can cover the demand better in places that the wind can't, which is a natural thing due to the fact that most energy consumption is done during sun hours, as we saw in the ponderation study in *Table 4-16*. Even though, the higher the value of solar coefficient (compared with other solar coefficients) the weakest is the sun overall production (because it needs a bigger installed power to cover the demand).

When comparing the storage values (*Table 6-19),* we can see that around the Iberian Peninsula the results are similar between them (all the storage values are in the range between 1 100 000 and 1 600 000 MWh) with one exception for when we have the sun in Coruña, there we can see that the values of the storage are a bit bigger than the average values around the Iberian Peninsula. This is due to the fact that, as we saw in the ponderation study when we compared Barcelona and Coruña data set combinations in *Table 4-16*; we saw that on the earlier hours of the day the sun of Coruña had a lower ponderation that the other averages suns around the Peninsula, making it resemble more like the Scottish sun, but not that pronounced. This will be taken into account when choosing the best set for a combined sun + wind generation plant between all the studied sets.

As for the storage values (*Table 6-19)*, they can be classified into two main ranges; those with a relative small value (around or under 1 600 000 MWh) and those which are bigger that the said value. This is an important first point to take into account because we know that the boundary conditions are the same so the only things that really affect the storage capacity are the solar and wind profiles. This means that the data sets that best adapt to the demand are those which manage to make the storage as small as possible. Therefore, we get our first conclusion of this part:

The smallest a storage value we get, the better its generation parameters adapt to the demand due to the fact that they will need less installed power to cover it.

An example of this behavior can be easily seen in the case of Scotland-Sahara and Sahara-Scotland; these two sets use the same wind coefficients (in mixed positions) so the only difference is the sun generation. We know for certain that both combinations of three values adapt to the demand, because that is how we programmed it to be. But what we can also see is that the obtained storage value is a bigger when having the sun in Scotland than in Sahara. This is mostly due to the fact that, the sun profile of Sahara is much better than the Scottish one and has a higher value, needing a less amount of installed power to arrive to the same point than Scotland.

This brings us to the second part of this discussion; how to differentiate similar storage values data sets and how do we choose the best one. Here comes into play that we are still minimizing storage, and this storage is determined through coefficients, which at the same time are determined by the sun and wind profiles. Therefore, the better the coefficients adapt to the demand the smaller value they will have.

Is at this point that we can take a look back at the results obtained in the ponderation study, as we can see there are two unusual scenarios:

- Sets that had a high overall ponderation value but have large storage values: These sets consists on any set that has Scotland or Coruña as sun power generator; the explanation is that if we look ant the additive generation plots that have Scotland and Coruña (*Plots 4-10 and 4-14)* as sun generators, we see that they both have very high wind values but small sun values, and thus camouflaging the lack of sun with wind in the overall demand coverage. This makes that when we use the sun for a storage system; we need a bigger coefficient, and thus a bigger storage value and a worse option to choose. This confirms the conclusion we got in the mean hourly ponderation study that said that **even though a good coverage can be achieved through powerful winds only, if we have to combine them with sun generation and that generation is weak, we will find ourselves with worse results than slightly smaller but better distributed solar and wind values.**

- Sets that had a small overall ponderation value but have small storage values: This is primarily the case of the sets that have Madrid as the wind only coefficient; as we saw in the overall coverage (*Plot 4-4)* Madrid's wind is fairly small, but tends to peak in the late hours of the day, where the demand is higher. That is why, when having to select a coefficient, even though it has a bigger value than other winds like Coruña or Scotland, which excel at wind values, it makes the overall storage smaller due to the better demand adaptation, therefore, the Madrid set, as a support of a powerful sun set, like Sahara, can help optimize the obtained storage better than other sets that may have higher wind values. This confirms the other conclusion we got in the mean hourly ponderation study that said that **even if we have a small overall coverage value, if we combine the different locations cleverly, we can still obtain good power demand coverage.**

Therefore we can obtain the following conclusion:

When facing similar storage values, those with the smallest coefficients will be the ones that adapt better to the demand, due to the fact that they need less installed power to cover it.

This has to be taken with a grain of salt, because there are cases that can make us doubt, for instance; if we were asked: What is better the Sahara-Scotland data set or Barcelona-Scotland data set? In this case if we looked at the storage values we would see that they are not that separate from each other. But when looking at the coefficients we see that the sun of Barcelona has a lower value than the sun from Sahara. This is due to the fact that the Saharan sun profile is higher on the morning hours and lower on the afternoon hours, as it can be seen in the *Plot 4-12* which is the opposite from the sun from Barcelona, and because the energy demand requires more energy in the afternoon , the Barcelonan sun gets a lower coefficient. In the case of the wind, even though they are both very similar, the Saharan wind has a better adaptation to the demand.

Therefore if we had to choose between one and the other, the correct answer would be to take into consideration other important factors that we haven't used here; factors like installation costs, financial subventions, legislative regulations or location.

But as for concrete, individual results, in the studied data sets, the combination of solar and wind power that gives us the smallest storage system, and therefore has the best demand adaptation is the combination is contested between Sahara-Madrid, Sahara-Scotland and Barcelona-Scotland, that have very similar storage values; but because Scotland has a better (meaning lower) wind coefficient and the overall storage system is smaller in Sahara-Scotland than Barcelona-Scotland, the best possible power generation installation is:

A 72 360 MW solar plant and 16 702 MW wind plant in Sahara plus a 10 784 MW wind power plant in Scotland with a total storage of 952 203 MWh.

If we are looking only on the Iberian Peninsula and disregard the overseas territories we have that the maximum storage values are bigger due to the fact that we have lost great solar and wind locations like Sahara and Scotland, so between the remaining combinations we have that the best option to minimize the storage is the Barcelona-Coruña set, therefore, the best possible generation between the studied data is:

A 66 155 MW solar plant and 24 450 MW wind plant in Barcelona plus a 20 071 MW wind power plant in Coruña with a total storage of 1 348 914 MWh.

6.4. Hourly storage study

The next step into this storage study is to see, once we have determined the storage value and coefficients for each combination of data sets, how it will affect to the overall energy production. Because even though we have chosen the coefficients and the storage that best adapts to the demand; there are bound to be times where the generation is over or under producing, due to the fact that we have chosen for the generation of ten years only one value of installed power for each energy source.

So in this second part we ill be looking at how many hours the energy generation is in a state of excess (with the storage at full capacity and generating a surplus of energy), in bounds (the storage is not fully charged or discharged) or in state of deficit (with the storage discharged and in need of more energy). Also, s we did before we will begin by showing the plots of the deficit, storage use and excess of each set and how many hours did that set had such values, after that we will draw some conclusions.

As to not be repetitive with the explanation of each plot, it will be done on the following paragraphs, and then we will only show the results in each data set.

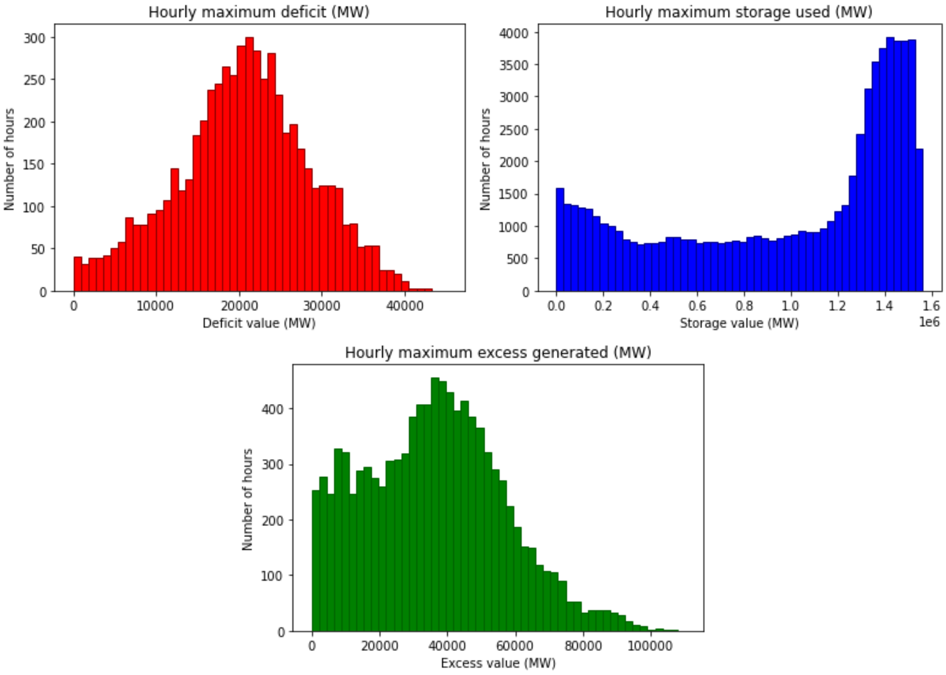
For all data sets we will be plotting three different histograms; the first one in red will represent the hourly maximum deficit, in here we will see how many hours and by how much there was an energy deficit with the parameters and storage chosen, an energy deficit is considered when the demand is higher than the generation for that hour and the storage is empty.

The second plot corresponds to the hourly maximum storage used, this graph corresponds to the number of hours that the storage is neither fully charged nor fully depleted; ideally, this section should have all the values, while the other two plots (deficit and excess) should be empty, that would mean that our storage and generation system is perfect, whit no excess of energy dumped nor energy deficit. But as we will see it will never happen, the motive of why will be discussed in the results.

Finally we have the third plot, this histogram will represent the number of hours that we are generating an excess of energy and, if nothing is done, that energy will be lost due to storage limitations.

One point worth mention before proceeding is the fact that the three plots can be seen as a continuation one from each other, the deficit plot's values are positive because they represent energy loss, while the other two are positive because they represent an energy gain. Thus the plots can be seen in the following order: First the deficit one (red) from right to left, then the storage one (blue) from left to right, and finally the excess one (green) also from left to right taking into account that the excess value corresponds to the energy generated over the maximum storage capacity (the sum of the histogram value plus the maximum storage value).

6.4.1. Barcelona-Madrid data set

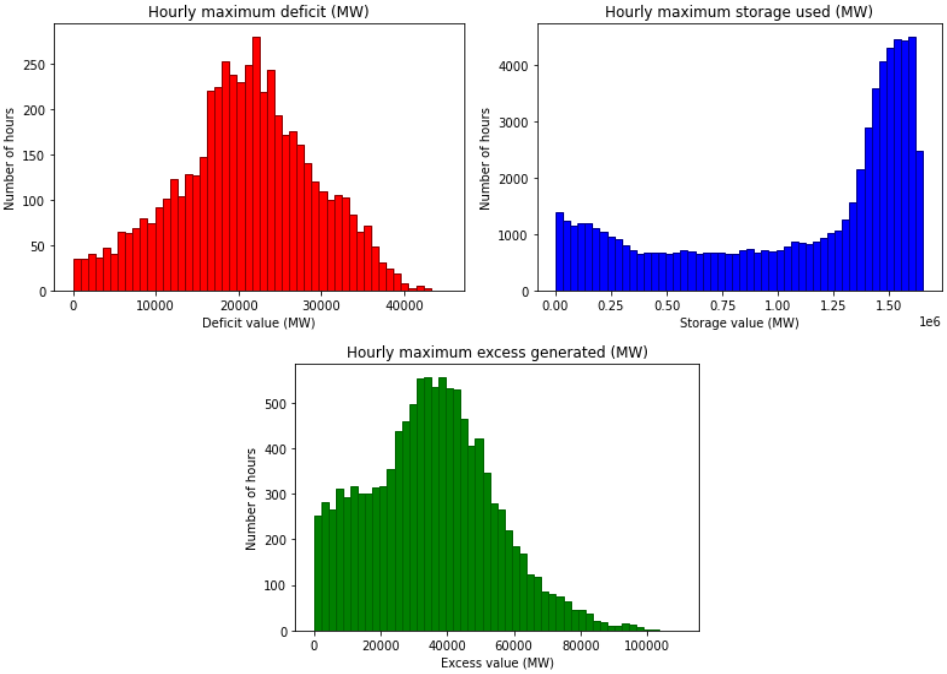

Hourly maximum deficit (MW)

Hourly maximum storage used (MW)

Hourly maximum excess generated (MW)

Plot 6-20, 6-21 and 6-22: Hourly maximum deficit, storage and excess plots for the data set

6.4.2. Madrid-Barcelona data set

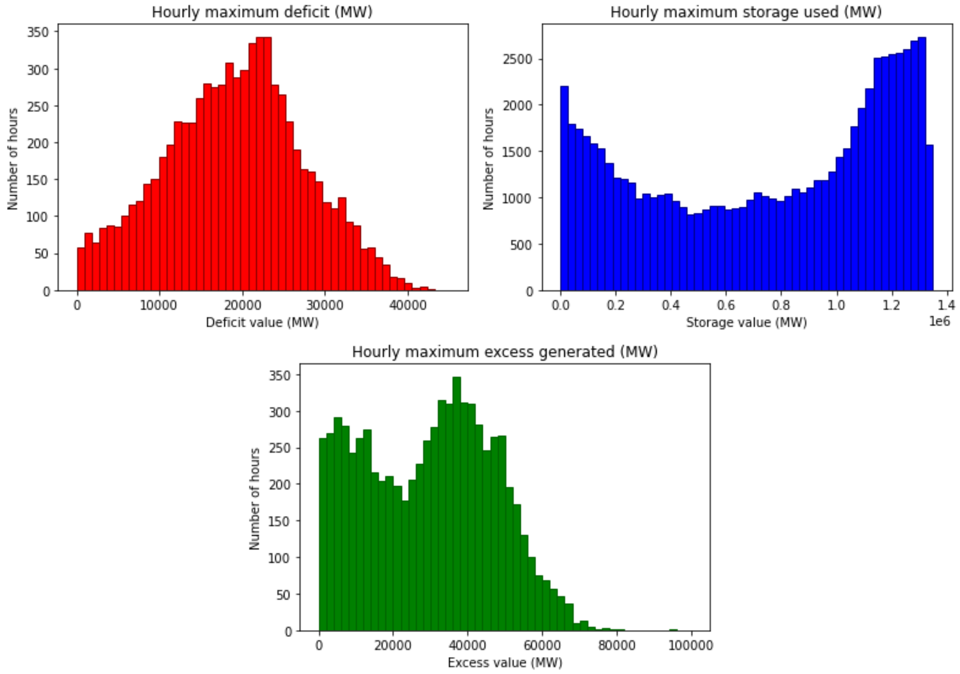

Hourly maximum deficit (MW)

Hourly maximum storage used (MW)

Hourly maximum excess generated (MW)

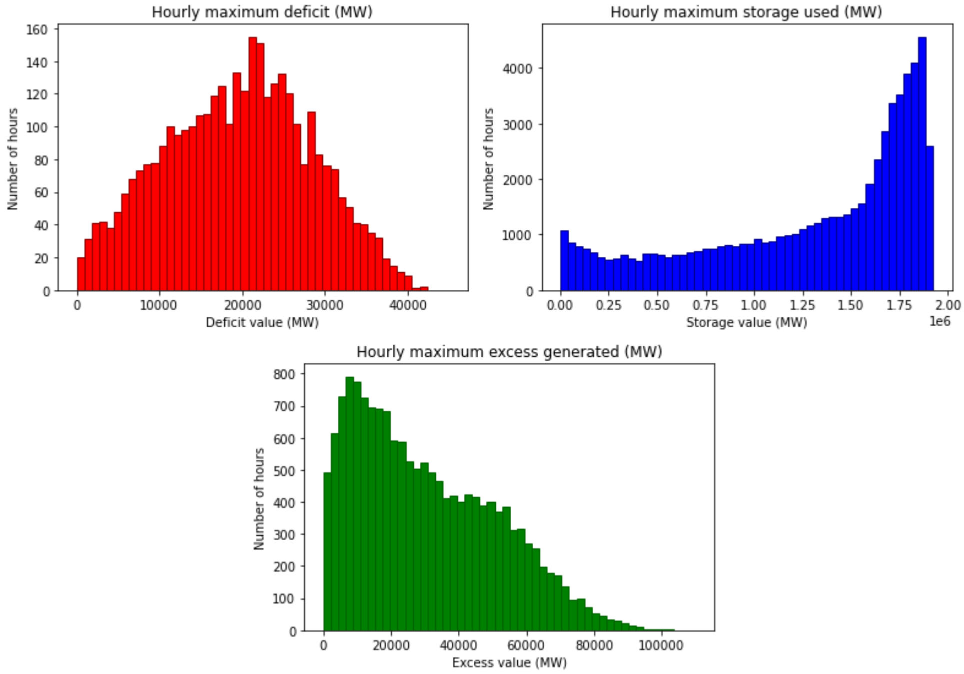
Plot 6-23, 6-24 and 6-25: Hourly maximum deficit, storage and excess plots for the data set

6.4.3. Barcelona-Coruña data set



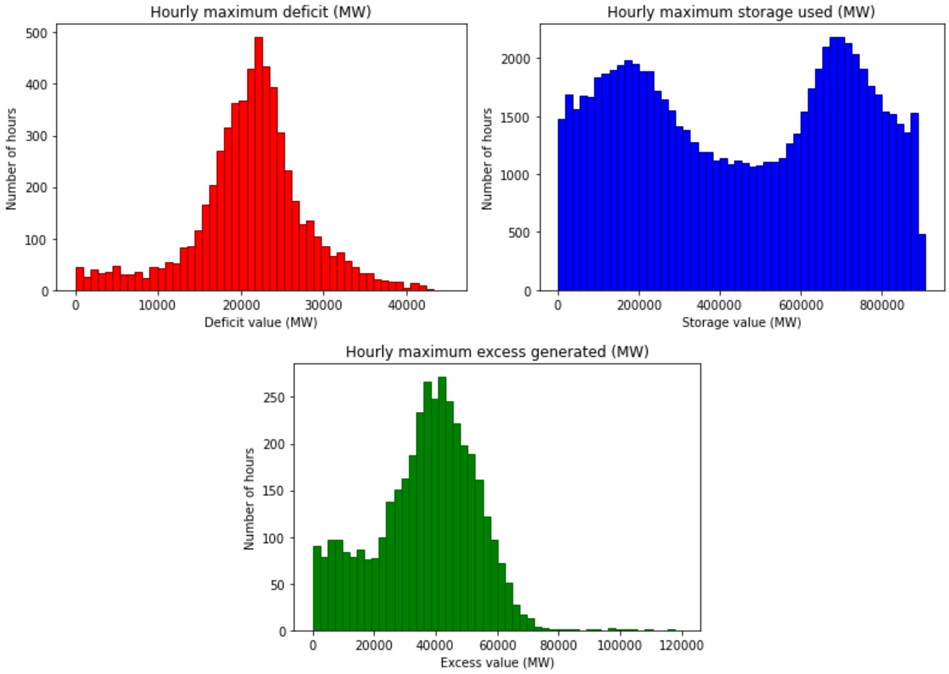
Plot 6-26, 6-27 and 6-28: Hourly maximum deficit, storage and excess plots for the data set

6.4.4. Coruña-Barcelona data set



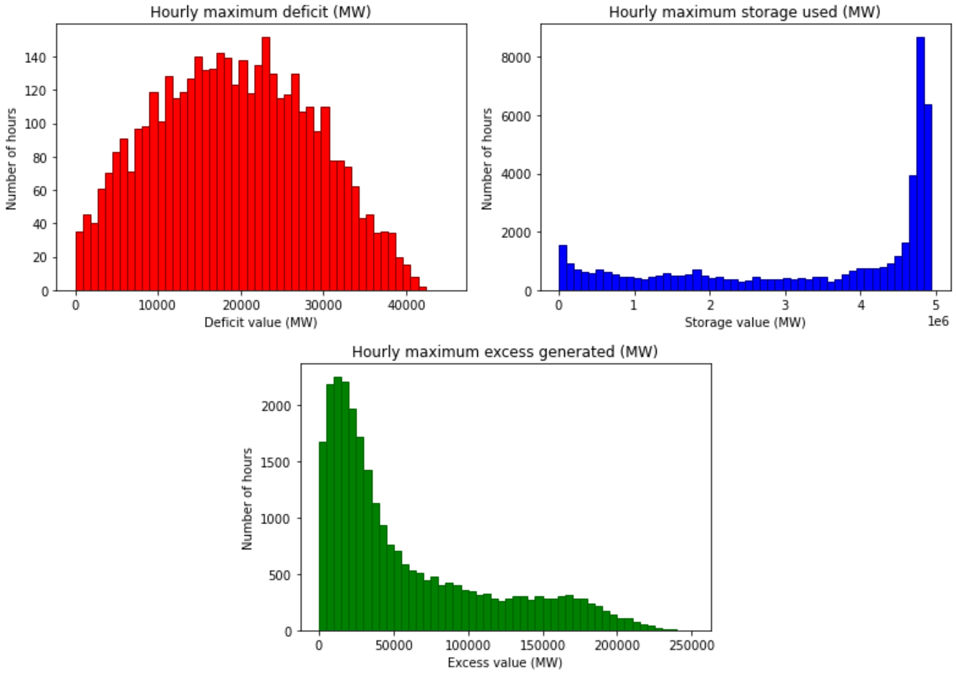
Plot 6-29, 6-30 and 6-31: Hourly maximum deficit, storage and excess plots for the data set

6.4.5. Sahara-Scotland data set



Plot 6-32, 6-33 and 6-34: Hourly maximum deficit, storage and excess plots for the data set

6.4.6. Scotland-Sahara data set



Plot 6-35, 6-36 and 6-37: Hourly maximum deficit, storage and excess plots for the data set

6.4.7. Results analysis

When focusing on the results, we can see that in every data set, in the *Plots 6-21, 6-24, 6-27, 6-30 and 6-36;* the histogram takes a basin form, which tends look more like an exponential the bigger is the storage calculated. The reason behind this is again the high coefficients that tend to keep the storage at full or near full capacity, giving us a higher number of excess periods and lower number of deficit periods.

We can also see that the maximum deficit value is always around 40 000 MW, which corresponds to the maximum demand; these deficits values are obtained when the demand is near to max and the production is almost zero for all the coefficients.

Even though this study is not necessarily linked to the decorrelation of the sun and wind itself, but rather to the coefficients obtained and how they do fill the storage, we can not extract as much information as in the previous section of this practical part. But what we can see is that the smaller the storage (due to a better adaptation to the demand because of the coefficients) leaves us with fewer excess time periods, which even though not undesirable, they are not optimal either. On the other hand, those smaller values make that we tend to have a higher number of deficit hours, as we clearly see in the Sahara and Scotland sets; the one with high storage values (Scotland-Sahara) has only around 600 hours of deficit, while the Sahara-Scotland has more. Still, this difference in deficit hours is compensated by the difference in excess hours, which gives us an overall better result in those sets that have a smaller storage, and therefore better adapted to the demand. This gives us another important conclusion:

A smaller storage system, which implies a better demand fitting, helps to optimize the deficit and excess hours of that said system.

6.5. Excess and deficit energy streaks

As an extra addition, and before proceeding to this second part's conclusions, we will talk briefly about the energy streaks; this is done because, as we said before it's important to know when we will be over or under energy covering capabilities and for how much long.

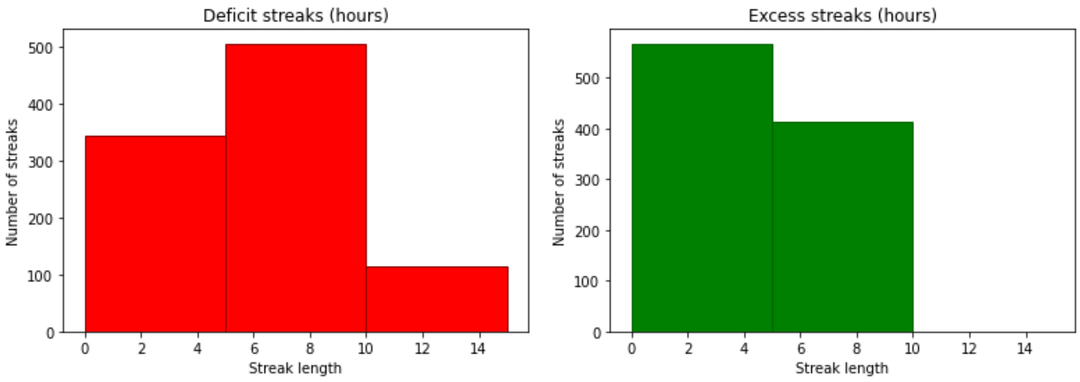
This part is mostly to check on the deficit streaks, duet to the fact that being prolonged periods of time without energy can be very negative; having an excess of generation for long periods of time doesn't seem that negative, but it means that the system is not optimal, which if we had used also prize for the optimization would be a negative aspect; nevertheless, because we haven't used it we will mainly focus on the deficit streaks.

The histograms show how many times a deficit or excess streak has happened, we define a streak as the number of hours that we find ourselves constantly over or under storage capabilities.

The next plots will show the deficit and excess streaks for the most extreme sets we have studied, Sahara-Scotland and Scotland-Sahara, so we will be able to see how the excess and deficit hours we saw in the previous section get distributed.

6.5.1. Sahara-Scotland data set

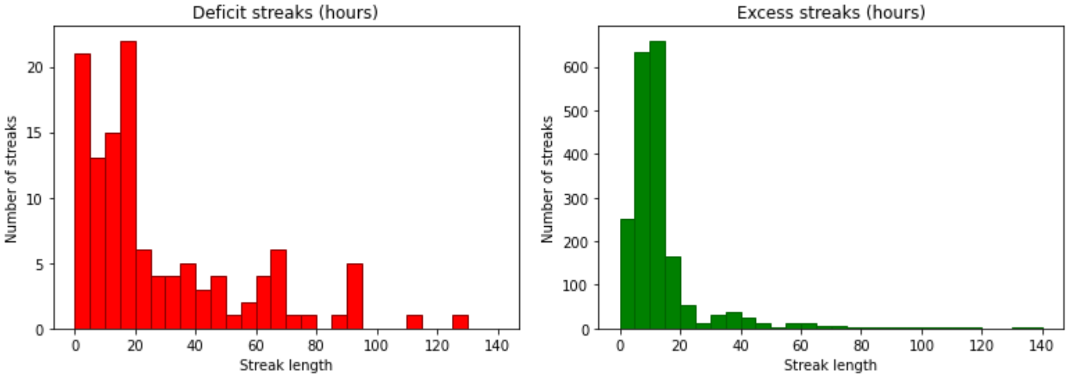
Here we can see the deficit and excess streaks for the Sahara-Scotland data set:



Plot 6-38 and 6-39: Histogram of the number deficit and excess streaks and their lengths

6.5.2. Scotland-Sahara data set

Here we can see the deficit and excess streaks for the Scotland-Sahara data set:



Plot 6-40 and 6-41: Histogram of the number deficit and excess streaks and their lengths

6.5.3. Results analysis

When comparing results what we can mainly see is that depending on the storage system the streaks can be more common but shorter, which happens for better combinations of data sets, as we can see in Sahara-Scotland (*Plots 6-38 and 6-39)*, even though we have a larger number of deficit streaks they are only at most 15 hours straight; while in the Scotland-Sahara set (*Plots 6-40 and 6-41)*, we have fewer streaks of different lengths but there are at least 17 streaks that last for more than 60 hours, which is generally worse because the shorter the streaks are, the more different options to cover them are (like batteries or engines for small scales and small combustion power plants for a more bigger scale)

With this we can also get the conclusion that:

Better storage systems tend to have a larger number of streaks than the worse ones, but the length of those streaks is shorter than the worse ones.

6.6. Overall conclusions of the storage study

In this second part we have thoroughly studied how the decorrelation of the sun and wind affects to a potential storage system created and have obtained some important conclusions that will be recapitulated here.

We have confirmed the following hypothesis introduced in the ponderation study:

Even though a good coverage can be achieved through powerful winds only, if we have to combine them with sun generation and that generation is weak, we will find ourselves with worse results than slightly smaller but better distributed solar and wind values.

Even if we have a small overall coverage value, if we combine the different locations cleverly, we can still obtain good power demand coverage.

Also, from the storage value and coefficient study we obtained that:

The smallest a storage value we get, the better its generation parameters adapt to the demand due to the fact that they will need less installed power to cover it.

When facing similar storage values, those with the smallest coefficients will be the ones that adapt better to the demand, due to the fact that they need less installed power to cover it.

From the studied data sets we found out that the best option was to install:

A 72 360 MW solar plant and 16 702 MW wind plant in Sahara plus a 10 784 MW wind power plant in Scotland with a total storage of 952 203 MWh.

Or if we are only looking at the Iberian Peninsula:

A 66 155 MW solar plant and 24 450 MW wind plant in Barcelona plus a 20 071 MW wind power plant in Coruña with a total storage of 1 348 914 MWh.

And from the hourly storage study and excess and deficit streaks we found out that:

A smaller storage system, which implies a better demand fitting, helps to optimize the deficit and excess hours of that said system.

And that:

Better storage systems tend to have a larger number of streaks than the worse ones, but the length of those streaks is shorter than the worse ones.

This has allowed us to understand better the concept of decorrelation and demand coverage and how we can take advantage from places which at first glance don’t seem as good as others for energy generation. And thus, we come to the final conclusion that **with the study of the energy demand from a certain region and the correct combination of decorrelated energy sources we will be able to determine the ideal locations for the energy generation.**

7. Future study lines

Having arrived to the end of the project, and before proceeding to the conclusions and the project's cost sections we will propose some future study lines that would be interesting to do in future projects. Some of these lines are the following:

- Optimizing the Python code so it can adapt to the physical parameters of the location when calculating the wind speed. This may not seem at first a huge deal, because the differences in speed the terrain rugosity provides is not that big, and in this project's case, the approximation was good enough. Having said that, the implementation of the rugosity factor and maybe some others that may affect also the wind or solar parameters may end up giving even more precise results.

- Extend the Python code so it can use other parameters in the storage calculation, like the type of storage that will be used or the total cost of the installation.

- Use the code for smaller scale demands, like cities or autonomous states, so we can get a more precise storage value that could be covered with an actual power installation in the selected locations. For example, we could be able to choose the best places from the province of Barcelona to install sun and wind power to cover all of Barcelona's demand.

A part from the points mentioned before, that basically are referencing the use of the code; another study line that we have come across along the duration of this project is the impact that the Sahara has had. We have not only seen that the Sahara has a very good sun bit also a great wind. This opens a lot of opportunities for energy generation, like for example filling the Great Erg with solar panels or wind turbines. This are, of course, just speculations and a lot of studies must be carried out to ensure the viability of this idea. Nevertheless the energy potential is out there and is a job for us humans to go and make use of it.

8. Budget

In this section the economical cost of this project will be calculated; here factors like the dedicated hours and the physical elements used during the project. In the following table we can see the breakdown of all the costs:

Concept	Quantity	Unitary cost	Total cost [€]
Personal expenses			**10,500 €**
Problem study	50 h	15 €/h	750 €
Information search	50 h	15 €/h	750 €
Code programming	100 h	15 €/h	1,500 €
Experiment simulations	50 h	15 €/h	750 €
Memory redaction	100 h	15 €/h	1,500 €
Hardware expenses			**1,150 €**
Laptop	1 unit	1000 €/ud	1,000 €
Computer peripherals	1 unit	150 €/ud	150 €
Software and licenses expenses			**294 €**
Microsoft Office license	1 unit	149 €/ud	149 €
Windows 10 license	1 unit	145 €/ud	145 €
Contingencies and other expenses			**500 €**
Contingencies and other expenses	-	-	500 €
Total (without IVA)			12,444 €
IVA (21%)			2,613 €
TOTAL COST			**15,057 €**

Table 8-1: Budget breakdown of the project.

As we can see the main cost corresponds to the human factor, going up to 10 500 € for the 350 hours poured in this project; the hourly wage has been chosen as the cost for all different sections of the project; 30 € each hour.

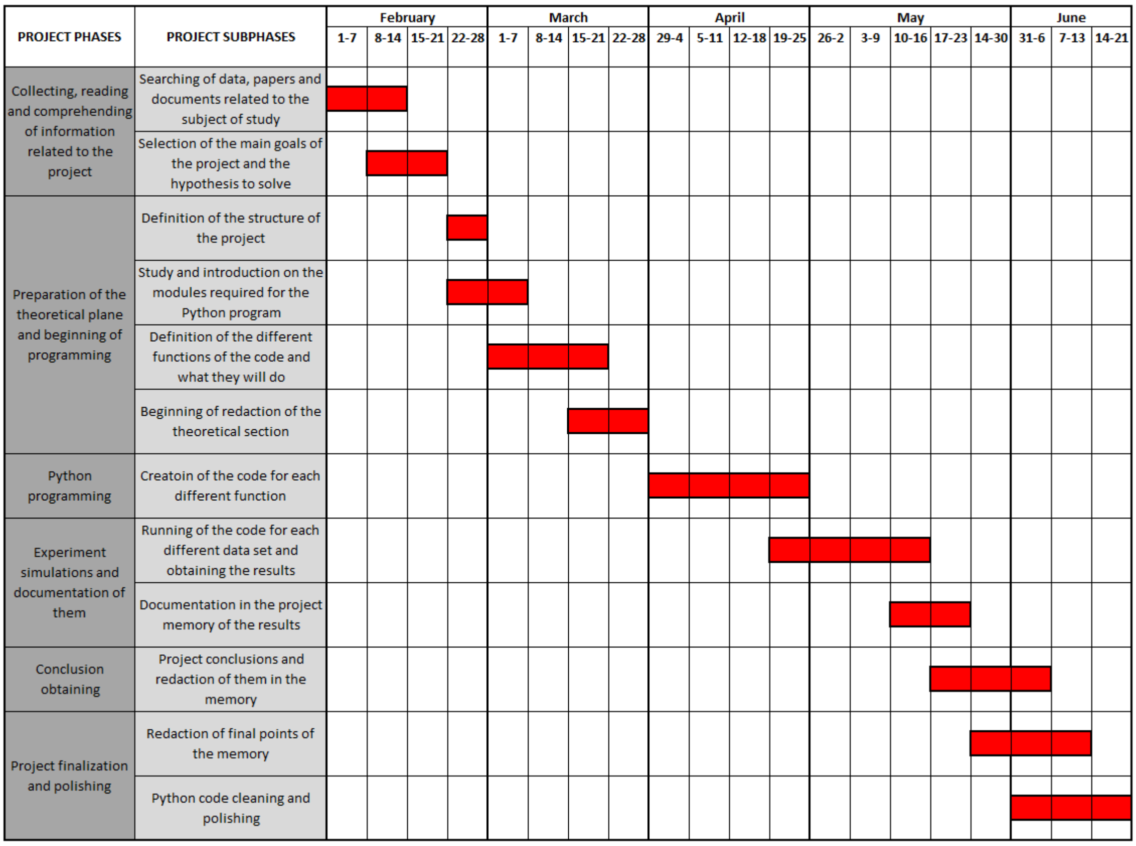
As for the software licenses, because Python is a free program, no expense is required into buying a license; as for the rest of the software, their license costs have been added to the total budget.

Finally, we have also added another section in the budget for contingencies and extra expenses; this section corresponds to all of those expenses that can't be classified as other type; for example travelling expenses or the internet and light used.

To sum up, this project's total cost ascends to 15 057 € (fifteen thousand fifty seven euros). As for future projects, the cost will be lower due to the fact that fewer hours will be required due to the experience and re using of work done here.

9. Project planning

In the following Gantt diagram we can see the different activities that have been done along the duration of this project:



PROJECT PHASES	PROJECT SUBPHASES	February				March				April				May					June		
		1-7	8-14	15-21	22-28	1-7	8-14	15-21	22-28	29-4	5-11	12-18	19-25	26-2	3-9	10-16	17-23	14-30	31-6	7-13	14-21
Collecting, reading and comprehending of information related to the project	Searching of data, papers and documents related to the subject of study	🟥	🟥																		
	Selection of the main goals of the project and the hypothesis to solve		🟥	🟥																	
Preparation of the theoretical plane and beginning of programming	Definition of the structure of the project				🟥																
	Study and introduction on the modules required for the Python program				🟥	🟥															
	Definition of the different functions of the code and what they will do					🟥	🟥	🟥													
	Beginning of redaction of the theoretical section							🟥	🟥												
Python programming	Creatoin of the code for each different function									🟥	🟥	🟥	🟥								
Experiment simulations and documentation of them	Running of the code for each different data set and obtaining the results												🟥	🟥	🟥	🟥					
	Documentation in the project memory of the results															🟥	🟥				
Conclusion obtaining	Project conclusions and redaction of them in the memory																🟥	🟥	🟥		
Project finalization and polishing	Redaction of final points of the memory																	🟥	🟥	🟥	
	Python code cleaning and polishing																		🟥	🟥	🟥

Plot 9-1: Gantt diagram of the different project phases

Conclusions

The aim of this project was to study the decorrelation between solar and wind power generations and how well would they cover an energy demand, in this case the Spanish one.

In order to do that, several data sets corresponding to different locations (both from the Iberian Peninsula and overseas) have been downloaded, edited and put against each other to create comparison sets, which eventually got compared with others from different locations.

Although the comparison method used on the ponderation study was relatively simple, we were able to obtain discerning results that separated worse location combinations from the better ones. These better locations were the ones which had higher overall values of sun and wind production and a higher decorrelation.

The highest demand coverages correspond to the sets corresponding to Coruña-Scotland and Barcelona-Scotland; with an overall demand coverage value of 46%; we can also see that the lowest values we obtain are around 31%. This means that with the studied sets; a good location helps covering more or less 15% more demand that another that may be less ideal.

Having said that, we also obtained some combinations that even though they gave us high demand coverage values it wasn’t because of their decorrelation between sun and wind, but just because their high wind values in both places.

This problem would be solved in the second part, when we did a quantitative study of the decorrelation. In there we defined a storage value and some installed power coefficients for the sun and each wind power parameter. In there we saw that the values which originally had high overall values for both sun and wind while also having a high decorrelation, had the minimum storage value and coefficients, meaning that they needed the lowest amount of inversion to cover the demand.

Meanwhile, the high values from the ponderation study that only depended on the wind power to cover the demand, saw themselves with relatively high storage values and high coefficients, due to having a very poor decorrelation.

This can be clearly seen if we compare the results from the Sahara-Scotland and Coruña-Scotland data sets. As we saw in the first practical part, the Coruña-Scotland data set has a higher demand coverage value than Sahara-Scotland, but when checking the storage values we see that their values are 1 977 123 and 952 203 MWh respectively; meaning that even though having a 2% less coverage than the Coruña-Scotland set; the Sahara-Scotland set needs only half the amount of storage to cover the demand; showing us how powerful a good sun and wind decorrelation can be.

After that we did an extra study on how the excess and deficit hours for the chosen storages behaved, and we saw that the best locations, and therefore the best storages, had the smallest amount of excess and deficit hours, and not only that their streaks were also the shortest ones.

Therefore, we arrived to the final conclusion that the decorrelation, or non simultaneity of the Sun and wind, and thus, solar and wind power have excellent demand covering capabilities; so the thought of this combination turning into an energy generation powerhouse is not only a very viable option to work as a base energy generation, but also could end up replacing all others when the proper improvements are done in their technology.

Acknowledgements

I would like to thank the direction of this project to my tutor, Vicente Cesar de Medina Iglesias, for his implication in the project, for being widely available for any doubt that I may have had and for granting me the opportunity to work in such an interesting and high relevance project.

I would also like to thank Jose Rebollo Pericot, who was the one who proposed me the topic of the project and has been deeply involved in its progress by giving new and interesting ideas; even though some of them, by different reasons have not been able to be seen in this project.

Finally, I would like to give my thanks to all my family and friends, who thanks to their unwavering support have helped me throughout this project.

References

Web references

- [1]: Vilaseró, Manuel; (January 22 2020); <España lidera el apagón mundial del carbón al cerrar plantas a un ritmo de vértigo>; accessed in June 5 2021. Source: https://www.elperiodico.com/es/medio-ambiente/20200122/espana-lidera-el-apagon-mundial-del-carbon-al-cerrar-plantas-a-un-ritmo-de-vertigo-7811565

- [2]: EIA.gov; (August 21 2020); <Natural gas explained. Factors affecting natural gas prices>; accessed in June 5 2021. Source: https://www.eia.gov/energyexplained/natural-gas/factors-affecting-natural-gas-prices.php

- [3]: Ministerio para la Transición Ecologica y el Reto Demográfico; (n.d.); <Centrales Nucleares en España>; accessed in June 7 2021. Source: https://energia.gob.es/nuclear/Centrales/Espana/Paginas/CentralesEspana.aspx

- [4]: Worldometer; (n.d.); <Coal Consumption by Country>; accessed in June 7 2021. Source: https://www.worldometers.info/coal/coal-consumption-by-country/

- [5]: World Nuclear Association; (n.d.); <Nuclear power in Italy>; accessed in June 8 2021. Source: https://world-nuclear.org/information-library/country-profiles/countries-g-n/italy.aspx

- [6]: Energy Sage; (n.d.); <Environmental impacts of hydropower>; accessed in June 8 2021. Source: https://www.energysage.com/about-clean-energy/hydropower/environmental-impacts-hydropower/

- [7]: Paul Denholm, Maureen Hand, Maddalena Jackson, and Sean Ong; (August 2009); <Land-Use Requirements of Modern Wind Power Plants in the United States>; accessed in June 9 2021. Source: https://www.nrel.gov/docs/fy09osti/45834.pdf

- [8]: Canadian Renewable Energy Network; (June 25 2002); <Solar Energy Technologies and Applications>; accessed in June 9 2021. Source: https://web.archive.org/web/20020625234427/http://www.canren.gc.ca/tech_appl/index.asp?CaId=5&PgId=302

- [9]: Jason Svarc; (March 10 2021); <Most Efficient Solar Panels 2021>; accessed in June 7 2021. Source: https://www.cleanenergyreviews.info/blog/most-efficient-solar-panels

- [10]: Kira Grogg; (2005); < Harvesting the Wind: The Physics of Wind Turbines>; accessed in June 10 2021. Source: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.589.2982&rep=rep1&type=pdf

- [11]: AN Bonus 600/44 Mk IV wind turbine; accessed in June 11 2021. Source: https://en.wind-turbine-models.com/turbines/690-an-bonus-600-44-mk-iv

- [12]: NEG Micon 1500/72 60 HZ wind turbine; accessed in June 11 2021. Source: https://en.wind-turbine-models.com/turbines/151-neg-micon-nm-72c-1500

- [13]: Nordex N80/2500 wind turbine; accessed in June 11 2021. Source: https://en.wind-turbine-models.com/turbines/89-nordex-n80-alpha
- [14]: Red Eléctrica Española; <Seguimiento de la demanda de energía eléctrica>; accessed in March 7 2021. Source: https://www.energysage.com/about-clean-energy/hydropower/environmental-impacts-hydropower/
- [15]: EU SCIENCE HUB; <Photovoltaic Geographical Information System>; accessed in February 15 2021. Source: https://re.jrc.ec.europa.eu/pvg_tools/es/#MR
- [16]: Giorgio Demurtas; (January 2011); <Vertical extrapolation of wind speed>; accessed in March 2 2021. Source: https://www.ingdemurtas.it/en/wind/vertical-extrapolation/
- [17]: Clarke Energy; (n.d.); <Energy Storage>; accessed in June 14 2 2021. Source: https://www.clarke-energy.com/energy-storage/

[18]: Ender Okandan; (n.d.); <Storage of Fossil Fuels (gaseous and Liquid)>; accessed in June 14 2021. Source: https://www.eolss.net/Sample-Chapters/C08/E3-14-05-02.pdf

Image references

- Plot 3-1: Data obtained from IEA.org. Source: https://www.iea.org/data-and-statistics/data-browser?country=WORLD&fuel=Energy%20supply&indicator=TPESbySource

- Plot 3-2: Data obtained from IEA.org. Source: https://www.iea.org/countries/spain

- Plot 3-3: Data obtained from IEA.org. Source: https://www.iea.org/countries/china

- Image 3-4: Obtained from inforse.org. Source: https://www.inforse.org/europe/dieret/Solar/solar.html

- Plot 3-5: Data obtained from solarpowereurope.org. Source: https://www.solarpowereurope.org/wp-content/uploads/2020/07/31-SPE-GMO-report-hr-hyperlinks.pdf?cf_id=34420

- Plot 3-6: Data obtained from irena.org. Source: https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2021/Apr/IRENA_RE_Capacity_Statistics_2021.pdf

- Image 3-7: Obtained from commons.wikimedia.org. Source: https://commons.wikimedia.org/wiki/File:Powercurve.png

- Plot 3-8: Data obtained from wind-turbine-models.com. Source corresponding to [11], [12], [13]

- Plot 4-3: Same data as 3-8 with the addition of a linear model shown in Equation 4-2

- Plot 4-4 to 4-15: Self done

- Table 4-16: Self done

- Tables and plots 6-1 to 6-41: Self done

- Table 8-1: Self done

- Table 9-1: Self done

- Annex 2: Sankey diagram: Obtained from IEA.org. Source: https://www.iea.org/sankey/

Annex

Annex 1: Python code

The code used in this project will be shown in this section. This same code has also been uploaded to GitHub and can be accessed via the following link: https://github.com/LuisBC123/Master-Thesis

```
"""
MASTER THESIS PYTHON CODE
by Luis Barrantes Coloma
"""
#%%
#The modules used on the program
import pandas as pd
import numpy as np
import os.path
import matplotlib.pyplot as plt
from scipy.optimize import minimize

#%%
def set_comparisons(lsets,sun_inst = 1, wind1_inst = 1, wind2_inst = 1, write = -1):
    """
    Main program for the first practical section:

    Given a list of sets, create the different set comparisons between them using tham as sun + wind and wind only (both cases);
    one set will not be compared with itself.
    """
    fw = open('Set comparisons.csv','w')    #File where we store the comparisons done
    dic_sets = {}                           #We also save them as a variable

    fw.write('Sun + wind set;Wind only set;Demand coverage\n')
    for i in range(len(lsets)):
        for j in range(len(lsets)):
            if i!=j:
                #We convert the downloaded sets to be able to work with them; creating a temporary file ending in _ready.csv
                set_i = df_structure(lsets[i]+'.csv')
                set_i_short = set_i[:-4]
                set_i = set_i[:-4]+'_ready.csv'
                set_j = df_structure(lsets[j]+'.csv')
                set_j_short = set_j[:-4]
                set_j = set_j[:-4]+'_ready.csv'

                #For each combination of sets we merge them into one set and do the different studies
                df = df_merge(set_i, set_j)
                filename = 'Comp-'+set_i_short +'-'+ set_j_short+'.csv'
                df2, dmed = df_edit (df.copy(),sun_inst, wind1_inst, wind2_inst, filename, write)
```

```
            dic_sets[filename[:-4]] = df2

            #We write in another file the overall demand coverage:
            fw.write(set_i_short+';'+set_j_short+';'+str(dmed)+'\n')
   fw.close()
   return dic_sets
#%%
def df_structure(set_name):
   """
   We transform the downloaded file into one we can work with.
   """
   #If the file we are going to create it already exists, we just use it
   if os.path.isfile(set_name[:-4]+'_ready.csv'):
      return set_name

   fr_set = open (set_name,'r')
   fr_demand = open ('Energy demand 2007-2016.csv','r')

   fw = open (set_name[:-4]+'_ready.csv','w')

   #We store the location data to know the exact parameters used in obtaining it; even though we
will not be using it in the project
   dparam = {}
   dparam['latitude'] = fr_set.readline().replace('"','').split()[3]
   dparam['longitude'] = fr_set.readline().replace('"','').split()[3]
   dparam['elevation'] = fr_set.readline().replace('"','').split()[2]
   fr_set.readline()
   fr_set.readline()
   fr_set.readline()
   dparam['slope'] = fr_set.readline().replace('"','').split()[1]
   dparam['azimuth'] = fr_set.readline().replace('"','').split()[1]
   dparam['nom_power'] = fr_set.readline().replace('"','').split()[8]
   dparam['system_losses'] = fr_set.readline().replace('"','').split()[3]

   fw.write('lat (º);'+ dparam['latitude'] + ';long (º);' + dparam['longitude'] + ';elev (m);' +
dparam['elevation'] + '\n')
   fw.write('slope (º);'+ dparam['slope'] + ';azimuth (º);' + dparam['azimuth'] + ';nom power (kW);'
+ dparam['nom_power'] + ';syst losses (%);' + dparam['system_losses'] +'\n\n')

   fw.write('time;P;WS10;WS120;Demand\n')
   fr_demand.readline()
   fr_set.readline()

   #We start merging all the 10 years data (sun + wind + demand)
   while True:
      l_set = fr_set.readline().strip('\n').split(',')
      l_dem = fr_demand.readline().strip('\n').split(';')

      if len(l_set) == 1: #After all data we have a blank space, then we exit the loop
         break
```

```
        dem = l_dem[1]

        #We calculate the wind speed at 120 meters; more information in the memory's section: 5.2.
Preprocessing of the project
        v120val = float(l_set[5]) * (np.log(120/0.1) / np.log(10/0.1))
        v120 = str(v120val)[:4]

        #We remove unnecessary values and add the demand to the individual set
        fw.write(l_dem[0]+';'+l_set[1]+';'+l_set[5]+';'+v120+';'+dem+'\n')

    fr_set.close()
    fr_demand.close()
    fw.close()
    return set_name

#%%
def df_merge(solar_set, wind_set):
    """
    We create the combination of sets (sun + wind and wind only),  more information in the
memory's section: 5.2. Preprocessing section.
    """
    set_name = solar_set.rstrip('ready.csv').strip('_')+'/'+wind_set.rstrip('ready.csv').strip('_')+':'
    df_sol = pd.read_csv(solar_set, sep = ';',skiprows = 3)
    df_wind = pd.read_csv(wind_set, sep = ';',skiprows = 3)

    del df_sol['WS10']
    df_sol.columns = [set_name+'time','Sun_power','Wind1_speed','Demand']
    df_wind = df_wind['WS120']
    df_sol.insert(3,'Wind2_speed',df_wind)
    return df_sol
```

```
#%%
def df_edit(df, sun_inst, wind1_inst, wind2_inst = 0, filename = -1, write = -1):
    """
    Uses the data obtained from df_merge and does the ponderation study; more information in the memory's section: 5.2. Preprocessing.
    """
    #If we only have 1 value of installed wind we set the two at the same value
    if wind2_inst == 0:
        wind2_inst = wind1_inst

    #Total installed sun power; we have it in W gen / kW installed; we convert it to MW gen / MW installed
    df['Sun_power'] = df['Sun_power'] / 1000 * sun_inst

    #Wind power, we calculate it using the power curve as mentioned in the memory's section: 5.2. Preprocessing
    df.insert(3,'Wind1_power',wind_calc(df['Wind1_speed'],wind1_inst))
    df.insert(5,'Wind2_power',wind_calc(df['Wind2_speed'],wind2_inst))

    #We create the ponderation columns, firts we need to obtain the maximum values of all generations and demand
    solar_max_power = df.loc[:]['Sun_power'].max()
    wind1_max_power = df.loc[:]['Wind1_power'].max()
    wind2_max_power = df.loc[:]['Wind2_power'].max()
    max_demand = df.loc[:]['Demand'].max()

    #We insert the ponderation columns; the generations scaled between 0 and 1 and the demand between 0 and 3; for more information consult the memory's section: 5.2. Preprocessing
    df.insert(2,'Sun_pond',df['Sun_power']/solar_max_power)
    df.insert(5,'Wind1_pond',df['Wind1_power']/wind1_max_power)
    df.insert(8,'Wind2_pond',df['Wind2_power']/wind2_max_power)
    df['Demand_pond'] = df['Demand']/max_demand*3

    #We proceed to calculate the total generation ponderation and demand coverage ponderation
    df['Gen_pond'] = df['Sun_pond'] + df['Wind1_pond'] + df['Wind2_pond']
    df['Dem_cov'] =  df['Gen_pond'] / df['Demand_pond']

    #Finally, we calculate the overall demand coverage
    Dem_cov_med = np.mean(df['Dem_cov'])

    #Write the dataframe in a .csv format, with a default name if not given any
    if write == 1:
        if not isinstance(filename, str):
            filename = 'set.csv'
        df.to_csv(filename,sep=';')

    return df, Dem_cov_med
```

```
#%%
def wind_calc(column,num_gen):
    """
    Auxiliary function of df_edit; calculates the wind power from the wind speed using the equation
described in the memory's section: 5.2. Preprocessing
    """
    l = []
    for v in column:
        if v < 3.5 or v > 25:
            l.append(0*num_gen)
        elif v <= 14:
            a = 0.0952 * v - 0.3333
            l.append(a*num_gen)
        else:
            l.append(1*num_gen)
    return l

#%%
def graph(df):
    """
    This function plots the graphs used in the hourly ponderation study
    """
    #We temporarily rename the columns for easier use
    df = df.rename(columns = {'Dem_cov': 'pond'})
    df = df.rename(columns = {df.columns[0]: 'time'})

    X = np.linspace(0,23, 24, endpoint=True)
    l_average = []
    l_average_sun = []
    l_average_wind1 = []
    l_average_wind2 = []
    l_dem = []

    #We calculate the average values for each hour of the 10 years
    for e in range(24):
        e = str(e)
        if len(e) == 1:
            e = '0'+ e
        l_average.append(df[df['time'].str[-2:]==e]['pond'].mean())
        l_average_sun.append((df[df['time'].str[-2:]==e]['Sun_pond'] * df[df['time'].str[-2:]==e]['pond']
/ df[df['time'].str[-2:]==e]['Gen_pond']).mean())
        l_average_wind1.append((df[df['time'].str[-2:]==e]['Wind1_pond'] * df[df['time'].str[-
2:]==e]['pond'] / df[df['time'].str[-2:]==e]['Gen_pond']).mean())
        l_average_wind2.append((df[df['time'].str[-2:]==e]['Wind2_pond'] * df[df['time'].str[-
2:]==e]['pond'] / df[df['time'].str[-2:]==e]['Gen_pond']).mean())
        l_dem.append(df[df['time'].str[-2:]==e]['Demand_pond'].mean())

    arr_average = np.array(l_average)
```

```
   #Additive generation source vs demand plot
    plt.figure()
    plt.plot(X, np.array(l_average_sun) , color = 'orange', linewidth = 1, linestyle = '-', label = 'Sun production')
    next_layer = np.array(l_average_sun) + np.array(l_average_wind1)
    plt.plot(X, next_layer , color = 'darkblue', linewidth = 1, linestyle = '-', label = 'Sun + Wind 1 production')
    next_layer +=  np.array(l_average_wind2)
    plt.plot(X, next_layer , color = 'cyan', linewidth = 1, linestyle = '-', label = 'Sun + Wind 1 + Wind 2 production')
    plt.plot(X, l_dem , color = 'red', linewidth = 1, linestyle = '-', label = 'Power demand')

    plt.xlim(0,23)
    plt.xticks(list(range(24)))
    plt.title('Additive generation source vs demand')
    plt.legend(loc = 'best')

    #Percenage dof demand covered plot
    plt.figure()
    plt.plot(X, arr_average , color = 'green', linewidth = 1, linestyle = '-', label = 'Percentage of demand covered')
    plt.xlim(0,23)
    plt.xticks(list(range(24)))
    plt.ylim(0,1)
    plt.title('Percentage of demand covered')
    plt.legend(loc = 'best')
```

```
#%%
def storage(d_frame,threshold,hours = 24):
    """
    The main code for the second practical section (the storage study); we use the same sets
created in the set_comparisons part
    """
    #We remove unnecessary columns
    df = d_frame.rename(columns = {d_frame.columns[0]: 'time'})
    df.pop('Sun_pond')
    df.pop('Wind1_speed')
    df.pop('Wind1_pond')
    df.pop('Wind2_speed')
    df.pop('Wind2_pond')
    df.pop('Demand_pond')
    df.pop('Gen_pond')
    df.pop('Dem_cov')

    #We calculate the installed power coefficients for each time period; for more information chech
the memory's section: 7.3. Storage values and coefficients calculation.
    max_thresh_sun = 1000000
    max_thresh_wind = 1000000
    b_sun = (0, max_thresh_sun)
    b_wind = (0,max_thresh_wind)
    bounds = (b_sun,b_wind,b_wind)
    con = {'type':'eq', 'fun': constraint}
    coef_0 = [10,10,10]
    daily_storage = []

    for d in range (int(df.shape[0]/(hours))):
        global df_slice
        df_slice =  df.iloc[d * hours: (d + 1) * hours]
        if df_slice.shape[0] != hours:
            break
        params = minimize(objective, coef_0, bounds = bounds, constraints = con)
        daily_storage.append([d,params.fun,params.x])


    df_storage = pd.DataFrame.from_records(daily_storage)
    coef_sun = 0
    coef_w1 = 0
    coef_w2 = 0

    for e in df_storage[2]:
        coef_sun += e[0]
        coef_w1 += e[1]
        coef_w2 += e[2]
    coef_sun /= df_storage.shape[0]
    coef_w1 /= df_storage.shape[0]
    coef_w2 /= df_storage.shape[0]
```

```
    chosen_coefs = [coef_sun,coef_w1,coef_w2]

    #With the obtained values we proceed to calculate the new storage values; where we will purposely leave a certain % out as a voluntary loss
    new_sto = []
    for d in range (int(df.shape[0]/(hours))):
        df_slice =  df.iloc[d * hours: (d + 1) * hours]
        if df_slice.shape[0] != hours:
            break
        new_sto.append(objective(chosen_coefs))
    new_sto = np.array(new_sto)
    max_storage = np.percentile(new_sto,threshold*100)
    values = chosen_coefs

    #Maximum storage each period of time plot (full plot)
    plt.figure ()
    N, bins, patches = plt.hist(new_sto, bins = [x*new_sto.max()/100 for x in range(101)], edgecolor = 'k')
    plt.title('Maximum storage each period of time (through regression)')
    plt.ylabel('Number of periods')
    plt.xlabel('Storage value (MW)')

    for i in range (len(patches)):
        if bins[i] >=  max_storage:
            patches[i].set_facecolor('k')
        else:
            patches[i].set_facecolor('b')

    #Maximum storage each period of time plot (after capping)
    plt.figure ()
    plt.hist(new_sto, color='blue', bins = [x*max_storage/100 for x in range(101)], edgecolor = 'k')
    plt.title('Maximum storage each period of time (through regression)')
    plt.ylabel('Number of periods')
    plt.xlabel('Storage value (MW)')

    #Here we take the maximum storage value keeping in mind the threshold and the maximum storage value
    df['New_generation(MW)'] = df['Sun_power'] * values[0] + df['Wind1_power'] * values[1] + df['Wind2_power'] * values[2]
    df['Hourly_deficit(MW)'] = (df['New_generation(MW)'] - df['Demand']) #Positive => Excess ; Negative => Defect
    df = storage_variation(df,max_storage)
    df['Storage_excess(MW)'] = df['Hourly_deficit(MW)']*(df['Storage_value(MW)']==max_storage)
    df['Storage_deficit(MW)'] = abs(df['Hourly_deficit(MW)']*(df['Storage_value(MW)']==0))

    #We plot the hourly storage study's 3 plots deficit, storage used and excess
    max_ex = int((df['Storage_excess(MW)'].max()//10000)+1) *10000
    max_def = int((df['Storage_deficit(MW)'].max()//5000)+1) * 5000
    df_aux = df[df['Storage_value(MW)'] < 0.99* df['Storage_value(MW)'].max()] #We crop tha 0 and maximum values as we consider them deficit or excess respectively
```

```
    df_aux2 = df_aux[df_aux['Storage_value(MW)'] > 0.1]

    plt.figure()
    plt.hist(df_aux2['Storage_value(MW)'], color='blue', bins = [x*max_storage/50 for x in
range(51)], edgecolor = 'darkblue')
    plt.title('Hourly maximum storage used (MW)')
    plt.ylabel('Number of hours')
    plt.xlabel('Storage value (MW)')

    plt.figure()
    plt.hist(df[df['Storage_excess(MW)']>0.1]['Storage_excess(MW)'], color='green', bins =
[x*max_ex/50 for x in range(51)], edgecolor = 'darkgreen')
    plt.title('Hourly maximum excess generated (MW)')
    plt.ylabel('Number of hours')
    plt.xlabel('Excess value (MW)')

    plt.figure()
    plt.hist(df[df['Storage_deficit(MW)']>0.1]['Storage_deficit(MW)'], color='red', bins =
[x*max_def/50 for x in range(51)], edgecolor = 'darkred')
    plt.title('Hourly maximum deficit (MW)')
    plt.ylabel('Number of hours')
    plt.xlabel('Deficit value (MW)')

    #We calculate the excess and deficit streaks
    l_deficit = []
    l_excess = []
    streak_start = ''
    streak = 0
    previous_day = 'N' #N = neutral; E=excess; D=Deficit

    for hour in range (int(df.shape[0])):
        if df['Storage_excess(MW)'][hour] != 0 : #If today is a Excess period
            if previous_day == 'E': #E streak continues
                streak += 1

            elif previous_day == 'D': #D streak ends, E streak starts in this order
                l_deficit.append([streak_start, streak])
                streak_start = df['time'][hour]
                streak = 1

            else: #E streak starts
                streak_start = df['time'][hour]
                streak = 1

            previous_day = 'E'

        elif df['Storage_deficit(MW)'][hour] != 0: #If today is a Deficit period
            #print(df['time'][hour] +' is D')
            if previous_day == 'D': #D streak continues
                streak += 1
```

```
        elif previous_day == 'E': #E streak ends, D streak starts in this order
          l_excess.append([streak_start, streak])
          streak_start = df['time'][hour]
          streak = 1

        else: #D streak starts
          streak_start = df['time'][hour]
          streak = 1

        previous_day = 'D'

      else:  #If today is a Neutral period
        if previous_day == 'D': #D streak ends
          l_deficit.append([streak_start, streak])

        elif previous_day == 'E': #E streak ends
          l_excess.append([streak_start, streak])

        # If previous day was Neutral we do nothing
        previous_day = 'N'

  if l_deficit == []: #In the case there is no deficit
    l_deficit.append(['NONE',0])

  if l_excess == []:  #In the case there is no excess
    l_excess.append(['NONE',0])

  #We plot the excess/deficit stereaks
  dfe = pd.DataFrame.from_records(l_excess)
  dfd = pd.DataFrame.from_records(l_deficit)
  max_x = ((max(dfe[1].max(),dfd[1].max())//10)+1) #We get the x ticks over the latest multiple of
five of the maximum

  plt.figure(11)
  bins = [x*5 for x in range(max_x)]
  plt.hist(dfe[1], color='green', bins = bins, edgecolor = 'darkgreen')
  plt.title('Excess streaks (hours)')
  plt.ylabel('Number of streaks')
  plt.xlabel('Streak length')

  plt.figure(12)
  bins = [x*5 for x in range(max_x)]
  plt.hist(dfd[1], color='red', bins = bins, edgecolor = 'darkred')
  plt.title('Deficit streaks (hours)')
  plt.ylabel('Number of streaks')
  plt.xlabel('Streak length')

  return (df_storage, df, new_sto ,max_storage, values, dfe, dfd)
```

```
#%%
"""
This three functions are the ones used in the minimization calculation; in orther they are the
objective function, the constraint used
and a function to do the accumulative sum in each row
"""

def objective (c):
    """
    The function to minimize (the maximum storage)
    """
    global df_slice
    gen = df_slice['Sun_power'] * c[0] + df_slice['Wind1_power'] * c[1] + df_slice['Wind2_power'] *
c[2]
    deficit = gen - df_slice['Demand']
    cum_deficit = abs(deficit.cumsum())
    return cum_deficit.max()

def constraint(c):
    """
    The constraint to make the daily avg generation equal to the daily avg demand
    """
    global df_slice
    return df_slice['Sun_power'].sum() * c[0] +df_slice['Wind1_power'].sum() * c[1] +
df_slice['Wind2_power'].sum() * c[2] - df_slice['Demand'].sum()

def storage_variation(df,storage):
    arr = []
    for i in range(df.shape[0]):
        if i == 0:
            arr.append( min(max(df['Hourly_deficit(MW)'][i],0),storage))
        else:
            new_storage = min(max(df['Hourly_deficit(MW)'][i]+arr[-1],0),storage)
            arr.append(new_storage)
    df['Storage_value(MW)'] = arr
    return df
```

```
#%%
"""
One example for a set study
"""
#We create all the comparisons of these sets, and create the .csv files
lsets = ['set_Barcelona','set_Madrid','set_Sahara','set_Scotland','set_Coruña']
print('Creating Sets...')
dic_sets = set_comparisons(lsets,sun_inst = 1, wind1_inst = 1, wind2_inst = 1, write = 1) #This also does the ponderation calculations

#We choose this set to study (Barcelona-Madrid)
lstud = ['Comp-' + lsets[0] + '-' + lsets[1]]
print(dic_sets[lstud])

#We create the plots used in the ponderation study
print('Creating Graphs...')
graph(dic_sets[lstud])

#We do the storage analysis
print('Storage analysis...')
df1, df2, new_sto,sto, values,dfe, dfd = storage(dic_sets[lstud], 0.75, hours = 7*24)
print('Storage: %f\nCoefficients: '%sto,values)
```

Annex 2: Sankey diagram

We plot here the Sankey diagram for the world energy consumption shown in section 3.1.