UCLouvain

École polytechnique de Louvain

The Impact of COVID-19 on Energy Systems

Analysis of the shock in electricity production in France in 2020

Author: **Nicolas CLAUX**
Supervisor: **Francesco CONTINO**
Readers: **Panagiotis VARELAS, Hervé JEANMART**
Academic year 2021–2022
Master [120] in Mechanical Engineering

Abstract

This thesis is focused on the impact of the COVID-19 pandemic on energy systems, and especially on the production of electricity in France in 2020.

This subject is directly linked with the field of electricity production (or load) forecasting. Such a forecasting can be complex to make because of events that can modify the trend of the electricity production : wars, stock market crashes, pandemics, terrorist attacks, etc. Although the COVID-19 pandemic is the subject of this thesis, the work done here can be applied for another type of event.

In this thesis, after a brief contextualization of the COVID-19 pandemic in France and of its impact on the electricity production, a general forecasting model is presented. This model is able to forecast the production for the whole year 2020 and it is then used in a dynamic system to include the shock due to the COVID-19. This is followed by a discussion of the method used and a qualitative analysis of this shock.

The results obtained and the analysis made can be added to the already existing works about the forecasting of economic shocks.

Keywords

Energy, Electriciy production, COVID-19, Forecasting, Shock Forecasting

Acknowledgments

First I would like to thank all the people that contributed to this thesis.

Then I would like to thank my supervisor, Francesco Contino, for his support and my mentor, Panagiotis Varelas, for his good advices and his availibility all over the year.

I also would like to thank the "Athoise UCL" for its great support during all the years I spent at the university.

I also would like to thank "La fraté" for all the good times we spent since the early childhood.

I especially would like to thank my father, Eric, for all he did for me during all the years I spent at the university.

I also would like to thank my girlfriend, Gwenn, for her advices and support during the writing of this thesis.

Finally I would like to thank all my friends who supported me for so long.

Contents

List of Figures

List of Tables

1 Introduction

The forecasting in the field of energy and in particular for the electricity has become crucial for several reasons. The forecast of electricity load[1] is used to make sure that the supply of electricity will be sufficient and to plan the required investments in new production units [1]. On the other hand, this forecast can lead to a forecast of the cost of electricity and these two forecasts are the basis of the decisions process in energy companies [2] [3].

Usually, these forecastings follow the same trend every year but there are sometimes events that change the world and thus that can completely disturb this trend. Here are some examples : The World Wars I and II, a financial crisis like the ones in 1929 or in 2007-2008, a terrorist attack like the 11th September, or more recently the war in Ukraine. All these events can impact the production, cost and consumption of energy all around the world. In this way, the COVID-19 pandemic is not an exception since the lockdowns imposed by different governments completely modified the life habits of most of the population.

Thus, it can be interesting to study the impact that these events can have on energy and the event chosen in this thesis is the COVID-19 pandemic. Since the field of energy is very wide, the focus is made on the production of electricity in France during 2020, the first year of the pandemic. Again, the production of electricity is a vast subject so a clarification of how it is approached is given just below. The main objective of this thesis is to develop a forecasting model able, not only to just forecast the electricity production, but also to include an event like the COVID-19 pandemic. Then, with such a model, some analysis can be done if the year 2020 happened differently : the COVID-

---

[1] Electricity load is the electricity demand.

19 arrives sooner or later in Europe, the duration of the lockdown period is modified, etc. This model and the obtained results are presented after a short contextualization of the pandemic, and are followed by a discussion of the method used.

The detailed plan of the rest of this thesis is the following : section 2 presents the context of the pandemic and the measures that impacted the electricity production and explains in more details the structure of the thesis. Section 3 presents a general forecasting model with its parameters, how to use it and its validation. Section 4 modify the model presented in section 3 to make it able to include the COVID-19. It is also in this section that the results of the study are presented. Section 5 discusses the method used to obtain these results, section 6 presents the limitations ans possible improvements and section 7 concludes this thesis.

2 Contextualization of the Thesis

2.1 Context of the Pandemic and Research Question

The general subject of this thesis was : "The Impact of COVID-19 pandemic on the Energy Systems. How the pandemic affected the different technologies and the decarbonization?". The aim of this section is to clarify the approach that will be made.

First of all, here is a little reminder of the last years. At the end of 2019 - beginning of 2020, the COVID-19 is discovered and begins to spread all around the world. In March 2020, it arrives in Europe and the different governments take quickly very strong measures to stop the pandemic : a complete lockdown which implies the closure of schools and stores for example.

These measures had a direct impact on the energy systems. Indeed, because most industries, stores and big buildings were closed, there was less electricity or heating consumption in them. However, these consumptions raised at home because people were very invited to stay home. There was also a bigger consumption in hospitals because more and more patients were admitted in them. An impact of these smaller or greater consumptions is a big shock in energy demand, shock which “is set to be the largest in the last 70 years” [4]. According to [4], the global energy demand dropped by 6% with respect to 2019. To have some figures which talks more than words, here are several differences between 2019 and 2020, from the total mean electricity generation from 16 European countries (figures taken from [4]) :

- Drop of 28% in generation of fossil electricity.
- Drop of 14% in nuclear electricity.
- Increase of 15% of renewable electricity.

These drops or increases are taken from April 2020 with respect to April 2019. At this period of the year, the total mean electricity generation had dropped by 9% (compared to the mean of the 5 previous years, so from 2015 to 2019).

Other impacts linked to the lockdowns are for example the CO_2 emissions that dropped a lot as did the price of oil because of everyone staying home, thus there was a stop of use of vehicles. Again, to present figures, CO_2 emissions dropped by a maximum of 26% [4]. A figure for the impact on the oil market is given in [5] : "the P.P.I.[2] for crude petroleum fell 71.0 percent from January to April".

These data show that a global analysis can be very difficult because of the differences between the different fields of energy : electricity, pollution, oil, etc. Thus there is a need of making choice of only one field of study and then doing the analysis on this field only.

But such an analysis on only one field can also be difficult because of the differences within it : for example for the electricity[3] there are big differences between industrial or commercial demand and residential or medical demand. There is also a difference between short or long term demand : the short term demand drops during a lockdown but the long term demand can be expected to be the same as before the lockdown when the government releases the measures.

---

[2] Producer Price Index.

[3] Because it is what will be studied in this thesis.

These reasons (short or long term, different types of technologies, different types of reactions depending on the environment) show that there is a need to clearly determine how the subject is studied : focus only on one type of technology, study the differences between the impact on the short or long term,...

In this thesis, the focus is thus made on the global electricity production in France during the year 2020 and how to represent the COVID-19 impact, i.e. can a model be developed to correctly modelize the shock due to an event like this ?

Before going further in this work, the section 2.2 shows some graphs of what has been explained in this section but for France, since it is the studied country and the section 2.3 presents the detailed structure of the thesis, related to the objectives to reach.

2.2 Illustration in France

After a brief contextualization, the aim of this section is to illustrate some figures that were presented in section 2.1 for France and for the whole 2020 year and not only until April. The data used to plot these graphs are taken from [6].

The figure 2.1 shows the evolution of the production of fossil and nuclear electricity in France in 2019 and 2020. Although the production in 2020 was already smaller than in 2019 at the beginning of the year, a strong decrease can be observed just after the beginning of March, so the exact period of the announcement of the first lockdown in France. The productions of the two years rejoin in November, so the production can, as explained in the previous section, be expected to recover a long time after the event that caused the shock.

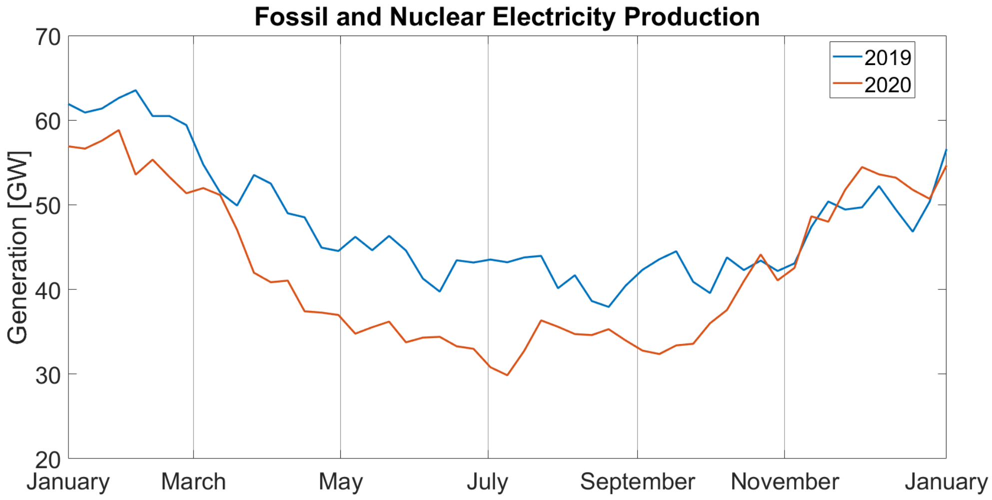


Figure 2.1: Evolution of the production of fossil and nuclear electricity in France in 2019 and 2020.

The figure 2.2 shows the evolution of the production of renewable electricity in France in 2019 and 2020. This graph is different from figure 2.1 because here the COVID-19 shock cannot be observed. The reason why is because the renewable electricity is not produced in function of the demand. Indeed, if there is sun, solar pannels will produce electricity no matter there is a high demand or not, the same happens for wind turbines with wind for example. Thus, as this electricity depends more on the weather than on the demand, the shock cannot be seen on a graph, and this is also the reason why this is the only type of electricity that was increasing[4] during the lockdowns : as all the other types were decreasing because of the decreasing demand, the share market of the renewable increased because its production remained constant because only affected by the weather.

---

[4] In share market.

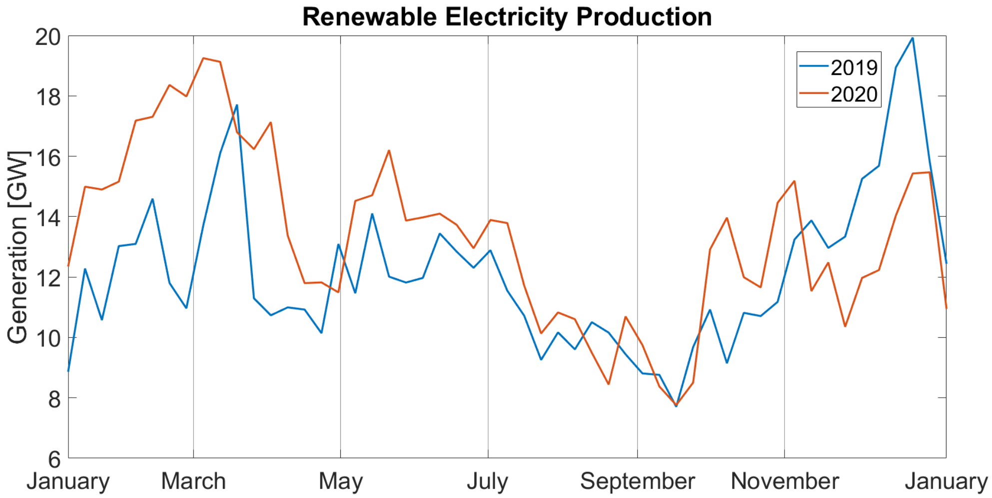


Figure 2.2: Evolution of the production of renewable electricity in France in 2019 and 2020.

The figure 2.3 shows the evolution of the global production of electricity in France in 2019 and 2020. This is maybe the more relevant graph because the shock can clearly be observed in March, with a decrease of the production and then a recovery around October-November.

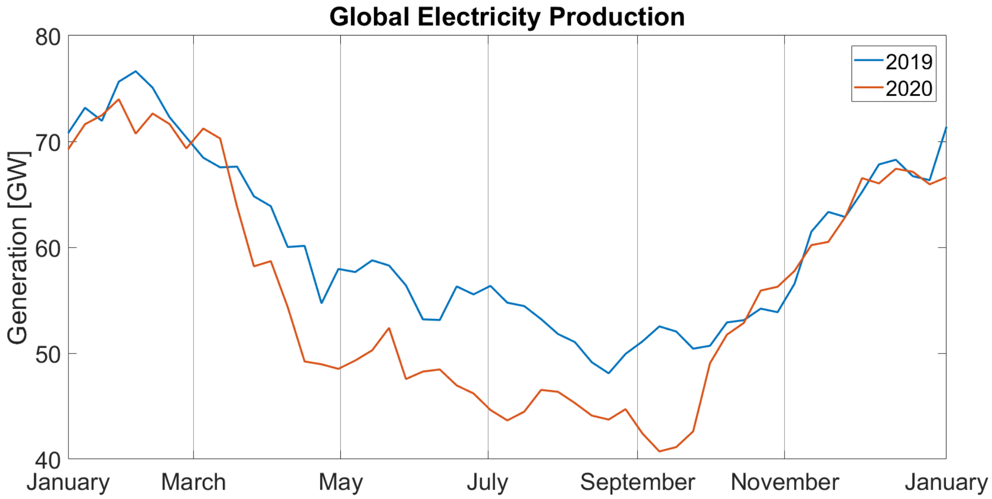


Figure 2.3: Evolution of the global production of electricity in France in 2019 and 2020.

2.3 Structure

As defined at the end of section 2.1, the aim of this thesis is to develop a model that can take into account a shock like the one due to the COVID-19 pandemic, or maybe another type of shock if it can be easliy translated into mathematical conditions.

First of all there is a need of a general forecasting model, which does not take COVID-19 into account but can forecast with high accuracy a whole year of electricity production (because the study of this thesis is on the 2020 year). This is what is done in section 3 : the author of the model is briefly presented, then the parameters and their modelization are explained in detail, followed by a brief presentation of the datasets needed and the performance evaluation, to conclude with the validation method which guarantees that the model works.

The reason why to first develop a general model which does not include COVID-19 is because first analysis can be done with this model and the forecasting of the 2020 year : if the model is accurate enough, the forecast is the expected production of the year if the shock does not exist, and so the comparison with the actual production can be more relevant than the comparison with the previous year. Another reason is because the general model will be used as an input in a model that can take a shock like COVID-19 into account. This is what is done in section 4 : after an introduction to determine how the COVID-19 pandemic can be mathematically represented and which models allow that representation, these models are presented with first a theoretical reminder and then their use in this thesis, and finally the modelization is explained as well as the validation method which, again, guarantees that the model works. Once this validation is done, some results are presented at the end of section 4.

After the presentation and validation of the model that includes COVID-19, there is still the question of the magnitude of a shock like this. This is the question studied in section 5 : how to determine the amplitude of a shock in function of its type ? (pandemic, war, stock market crash,...).

At the end of all these sections, limits, possible improvements and the conclusion of the thesis are presented.

3 General Forecasting Model

First of all, there is a need of a general forecasting model in order to compare the electricity production during the COVID-19 pandemic and if the year was normal. Once this is done, the question of how to include the pandemic into the model can be studied to analyse the impact of COVID-19 in different situations.

Thus this section presents[5] the general forecasting model, i.e. a model to predict normal electricity production. First the author and the origins of the model are briefly presented. Then the parameters included in the model are explained and the way that they can be modelized is detailed. Afterwards the datasets needed are presented . Finally, the performance evaluation[6] and the validation of the model close this section.

---

[5] In this presentation, the electricity load or electricity demand is used but in fact the data that will effectively be used is the production, because the available data are those of production and not those of electricity demand. This is not really a probeme because the production and the demand are expected to follow the same trend, and the section 3.6 will show that the model work this way. So when the term load is used in all this section, it is because the model is based on the load at its origin but in this thesis, the load refers in fact to the production.

[6] Since it is the aim of the model : accuracy.

3.1 Origins of the Model

In 2010, Dr. Tao Hong studied in [7] many different methods of load forecasting and in particular several methods based on multiple linear regressions and one model stood out from the crowd : the famous[7] Tao's Vanilla Benchmark model.

This model is based on the fact that the load depends on the weather (especially the temperature) and on calendar variables (the current time of the year influences the current load). Because of its ease of use and its good accuracy, it is a widely known model in terms of load forecasting.

In 2012 and 2014, Dr. Tao Hong led a team which organized 2 competitions of forecasting : the Global Energy Forecasting Competition 2012 (GEFCom2012) and the Global Energy Forecasting Competition 2014 (GEFCom2014) [8]. To visualize the aim of these competitions, here is the statement of the one of 2012 : "The topic for the load forecasting track is a hierarchical load forecasting problem: backcasting and forecasting hourly loads (in kW) for a US utility with 20 zones. The participants are required to backcast and forecast at both zonal level (20 series) and system (sum of the 20 zonal level series) level, totally 21 series." [9]. In 2014, the benchmark model was the reference used to rank the candidates [10].

Since the accuracy and ease of use of this benchmark model already have been proved in the literature, this is the model that will be used in this thesis. The following subsections explain in detail which parameters are involved and how to use and validate it.

---

[7] In the field of load forecasting.

3.2 Model Presentation

Section 3.1 presented a small litterature review of the used model and its results. The aim of this section is to present in detail the model and what are the big ideas behind its development.

The model can be written as :

$$\begin{aligned} Load_t \;=\; & \beta_0 + \beta_1\, Trend_t + \beta_2\, Hour_t + \beta_3\, Day_t + \beta_4\, Month_t + \beta_5\, Hour_t \times Day_t + \\ & \beta_6\, Temp_t \;+\; \beta_7\, Temp_t^2 \;+\; \beta_8\, Temp_t^3 + \\ & \beta_9\, Hour_t \;\times\; Temp_t \;+\; \beta_{10}\, Hour_t \;\times\; Temp_t^2 \;+\; \beta_{11}\, Hour_t \;\times\; Temp_t^3 + \\ & \beta_{12}\, Month_t \;\times\; Temp_t \;+\; \beta_{13}\, Month_t \;\times\; Temp_t^2 \;+\; \beta_{14}\, Month_t \times\; Temp_t^3 \end{aligned} \quad (3.1)$$

In equation 3.1, the subscript t is for the hour studied. For the other parameters, *Load* represents the electricity load (observed or forecasted), *Trend* represents the production trend (is the production increasing or decreasing, or is it approximately constant), *Hour*, *Day*, and *Month* are calendar variables which represents the period of the year and *Temp* represents the temperature. The idea is to use these data and parameters to compute the β_i coefficients which will forecast the future load with future data.

The big idea behind the equation 3.1 is that the load can be modeled as a function of calendar variables and temperature and interactions between several parameters. Each term is developed in details in sections 3.2.1 and 3.2.2. The data used to plot the graphs are those from France (since it is the country studied in this thesis).

3.2.1 Main Effects

1. **Trend** : The load follows a trend (either increasing, decreasing or even constant). The figure 3.1 shows that the production in France is essentially constant from 2015 to 2018 but there is a little decrease in 2019.

2. **Hour** : The load depends on the hour of the day. Indeed one can have the intuition that there is more demand at 16 p.m. than during the night. This is confirmed by figure 3.2, which shows the electricity production on the first Monday[8] of 2015. The figure 3.2 shows that there is a first peak from 8 a.m. to 14 p.m. before a greater peak around 20 p.m. . The figure 3.3 shows that the highest peak is not the same for all days : for some it is in the afternoon and for others it is in the evening. These peaks correspond to the moments when people are at work and when people are at home using electrical devices (kitchen, heating, lighting), so the moments when the production has to be more important.

3. **Day** : The load depends on the day of the week. Indeed a greater load is expected during the working days than during the week-end. This is confirmed by the figure 3.3 which shows the electricity production from 05/01/2015 to 11/01/2015[9]. Indeed, the figure 3.3 shows a production essentially constant from Tuesday to Friday but there is a big drop during the week-end.

4. **Month** : The load depends on the month of the year. This can be different in function of the studied country : hot countries should use more electricity during summer than cold countries because of the cooling devices. The figure 3.4 shows

---

[8] Since the first January of 2015 is a Thursday, the first interessant day is Monday because of the bank holiday on the first January and the week-end directly following.

[9] This is the week following the first Monday of 2015, for the same reason than the figure 3.2 shows the first Monday instead of the first of January.

the elecricity production during the year 2015 and shows that the peaks are during winter. This is because in France (and in Belgium) summer is not too hot and thus the peaks are when there is a huge need of heating.

5. **Temp** : The load depends on the temperature. Indeed, as explained just above, the highest production peak in France is during winter because of heating. So one can guess than temperature has an impact on the load : in general if weather is cold there will be more heating consumption and if weather is hot there will be more cooling consumption. The figure 3.5 shows the electricity production in function of the temperature from 2015 to 2019 and one can see that the curve can be approached by a 3^{rd} order polynomial of temperature.

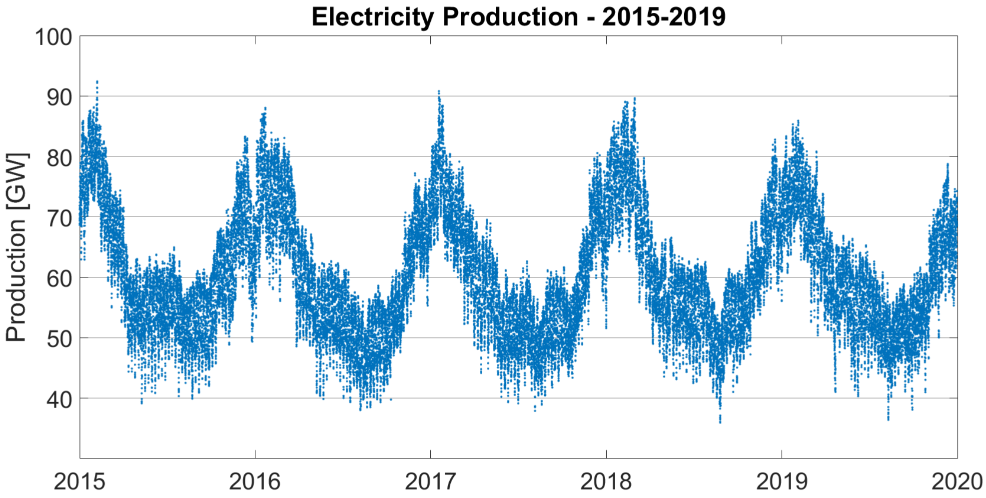


Figure 3.1: Electricity production in France from 01/01/2015 to 31/12/2019.

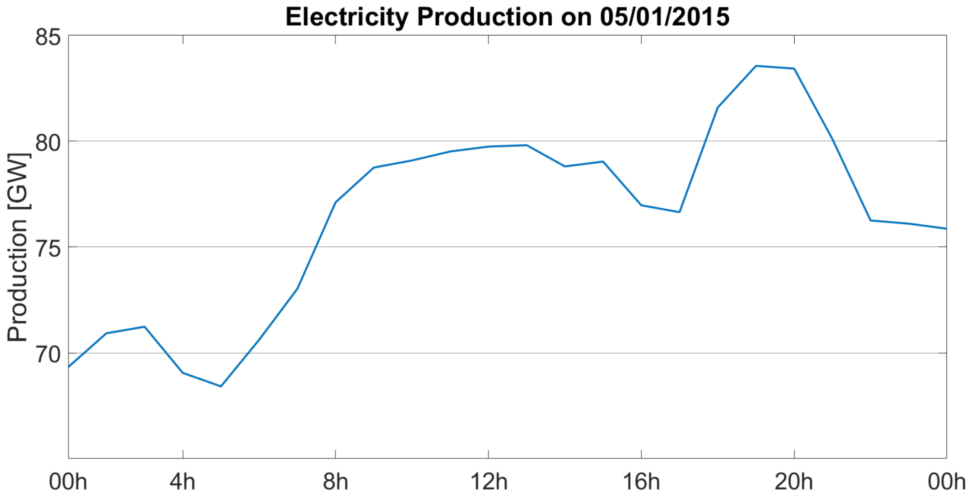


Figure 3.2: Electricity production in France on 05/01/2015.

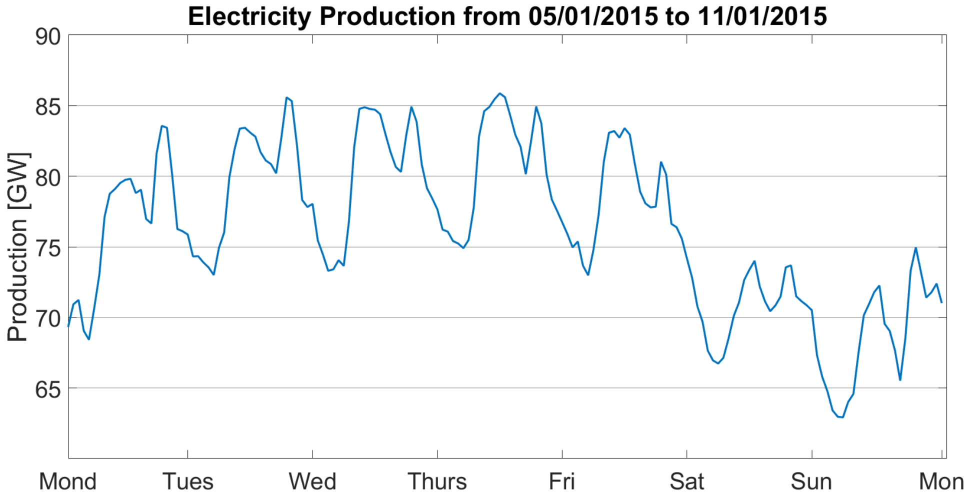


Figure 3.3: Electricity production in France from 05/01/2015 to 11/01/2015.

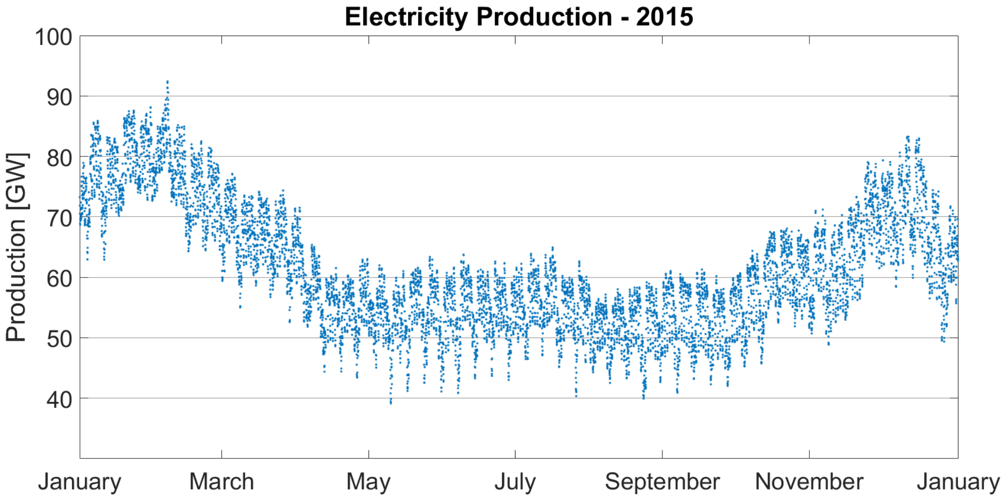


Figure 3.4: Electricity production in France from 01/01/2015 to 31/12/2015.

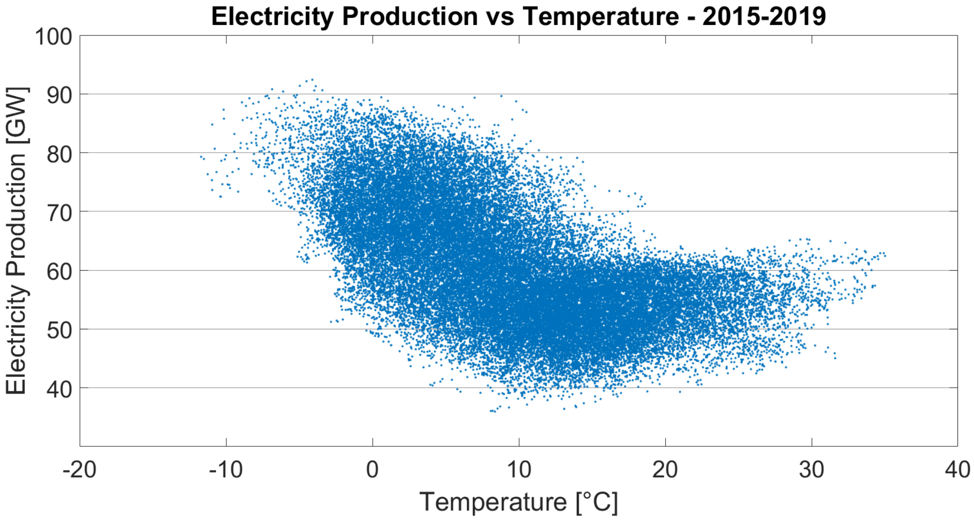


Figure 3.5: Electricity production vs Temperature in France from 01/01/2015 to 31/12/2019.

3.2.2 Interaction Effects

Now that the main parameters are defined and justified, one can notice that the model also takes into account the interactions that several parameters can have between them. The choice of these interactions is explained below.

1. **Hour** × **Day** : As already explained, the hour of the day and the day of the week influence directly the load but they are not independent. Indeed, the load should not be expected to be the same at 8 a.m. during working days than during week-ends (because people usually get up later during week-ends). This is shown on figure 3.3 : production during week-end mornings is much smaller than working days mornings (about 10 GW of difference so about 15% of difference). This is why the interaction between the hour of the day and the day of the week has to be included in the model.

2. **Hour** × **Temp** : As already explained, the hour of the day and the temperature influence directly the load but they are neither independent. Indeed, the temperature is expected to be smaller during the night than during the day[10]. So the interaction between the hour of the day and the temperature also has to be included in the model.

3. **Month** × **Temp** : As already explained, the month of the year and the temperature influence directly the load but they are neither independent. Indeed, the temperature is expected to be higher in August than in January for example, as shown on the figure 3.6. The highest peaks have already been identified during winter, so with the fact that temperature influences the load and that temperature is smaller during winter than during the rest of the year, the interaction

---

[10] There is no figure to illustrate that because it is evident.

between the month of the year and the temperature also has to be included in the model.

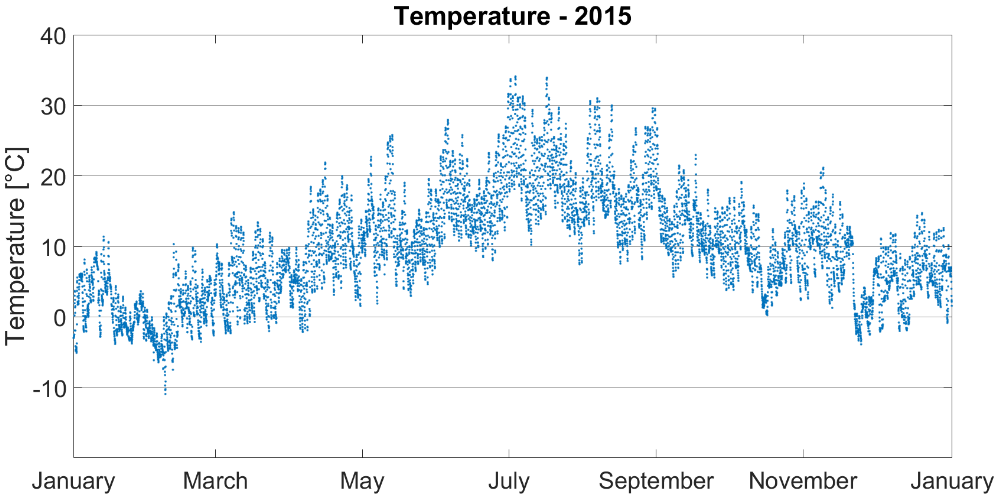


Figure 3.6: Temperature in France from 01/01/2015 to 31/12/2015.

3.3 Parameters Modelization

Section 3.2 presented the model and his parameters and explained why these parameters and their interactions have to be in the model. The aim of this section is to explain how to model these different parameters. Indeed, they must not be modeled the same way because they are not the same class of variables : $Trend$ and $Temp$ are quantitative variables but Day, $Hour$ and $Month$ are qualitative variables. The modelling of quantitative variables is easy but this is not always the case for qualitative variables for which there can be different methods of modeling.

3.3.1 Quantitative Variables

Quantitative variables (also called numerical variables) are easy to interpret and to model. The temperature in equation 3.1 is a very good example : there is nothing special to do in order to model it, the different values of temperature in the dataset can be directly included in the model. The two quantitative variable of the model are the trend and the temperature and their modelling is explained below.

1. **Trend** : The trend, as done in [11], is modeled by assigning to this variable an increasing[11] natural number for all the studied period. So $Trend$ value is 1 for the first hour of the studied period, 2 for the second and so on.
2. **Temperature** : The temperature is simply modeled by assigning to this variable the actual value of temperature from the dataset.

---

[11] It would be the same with a decrasing number, the β_1 coefficient would just be of the opposite sign.

3.3.2 Qualitative Variables

Qualitative variables (also called categorical variables) are not as easy to interpret and model as quantitative ones. Indeed, although the effects of hour, day and month on load have been explained in section 3.2, one can wonder how to represent the fact that the studied day is a Monday or a Wednesday for example. The three categorical variables of the model are the hour of the day, the day of the week and the month of the year and two methods of modeling them are presented below.

1. **Convert into number** : The idea (taken from [12]) of this method is to convert the categorical variables into numbers : the hours could be represented as natural number from 1 for the first hour of the day (00h-1h) to 24 for the last hour of the day (23h-00h), the days could go from 1 for Monday to 7 for Sunday and the months could go from 1 for January to 12 for December.
2. **Dummy variables** : The idea (see [13]) of a dummy variable is to define wether or not the condition of the variable is fullfilled : for example for the days, the condition can be "is the day Monday or not ?". Then a value of 1 is assigned to the variable is the condition is fullfilled and 0 is assigned if not. If there is just to determine wether the day is Monday or not, it is easy, the dummy variable can only take two values, but in the model there are 7 days so more conditions to evaluate.

Now that two methods have been briefly presented, the method of dummy variables is chosen and explained in more details. Let's keep the example of the condition "is the day Monday or not ?". So in the equation 3.1, the term involving the day can be written as :

$$\beta_3 \ \times \ Day_t = \begin{cases} \beta_3 \ \times \ 1 \ = \ \beta_3, & \text{if the day is Monday.} \\ \\ \beta_3 \ \times \ 0 \ = \ 0, & \text{if not.} \end{cases} \tag{3.2}$$

In equation 3.2, the only considered condition is wether the day is Monday or not but in the model, the condition to evaluate is "which is the current day ?", so the equation can be extended the equation 3.2 to the case which evaluates all days :

$$\beta_{3,i} \ \times \ Day_{i,t} = \begin{cases} \beta_{3,1} \ \times \ 1 \ = \ \beta_{3,1}, & \text{if the day is Monday.} \\ \\ \beta_{3,1} \ \times \ 0 \ = \ 0, & \text{if not.} \\ \\ \beta_{3,2} \ \times \ 1 \ = \ \beta_{3,2}, & \text{if the day is Tuesday.} \\ \\ \beta_{3,2} \ \times \ 0 \ = \ 0, & \text{if not.} \\ \\ ... & ... \\ \\ \beta_{3,6} \ \times \ 1 \ = \ \beta_{3,6}, & \text{if the day is Saturday.} \\ \\ \beta_{3,6} \ \times \ 0 \ = \ 0, & \text{if not.} \end{cases} \tag{3.3}$$

In equation 3.3, one can notice that there are only 6 evaluations of the condition "which is the current day ?". This is because if the day is Sunday, then all the others evaluations have given 0 and so there is no need to put a 7^{th} evaluation because it would be redundant : if all other variables are 0, then the day is automatically Sunday.

	$Day_{1,t}$	$Day_{2,t}$	$Day_{3,t}$	$Day_{4,t}$	$Day_{5,t}$	$Day_{6,t}$
Monday	1	0	0	0	0	0
Tuesday	0	1	0	0	0	0
Wednesday	0	0	1	0	0	0
Thursday	0	0	0	1	0	0
Friday	0	0	0	0	1	0
Saturday	0	0	0	0	0	1
Sunday	0	0	0	0	0	0

Table 3.1: Dummy variables for the parameter Day.

The table 3.1 resumes the use of the dummy variables for the parameter Day_t in the model. The same method is used for $Hour$ and $Month$.

One can notice that with this method and with the interactions effects between the temperature and the hours or the months, there is no need to keep the terms including only the temperature in the model because they are already included with the interaction with the dummy variables. The same is observed for the interaction between hours and days : thanks to the interation effect, the terms including only the hours or only the days can be removed because they are already included into the interaction term.

Finally the equation 3.1 can be rewritten as :

$$\begin{aligned}Load_t = \beta_0 + \beta_1\, Trend_t + \beta_{4,i}\, Month_{i,t} + \beta_{5,i,j}\, Hour_{i,t} \times Day_{j,t} + \\ \beta_{9,i}\, Hour_{i,t} \times Temp_t + \beta_{10,i}\, Hour_{i,t} \times Temp_t^2 + \beta_{11,i}\, Hour_{i,t} \times Temp_t^3 + \\ \beta_{12,i}\, Month_{i,t} \times Temp_t + \beta_{13,i}\, Month_{i,t} \times Temp_t^2 + \beta_{14,i}\, Month_{i,t} \times Temp_t^3\end{aligned} \tag{3.4}$$

3.4 Datasets

While sections 3.2 and 3.3 presented the model and how to use the differents parameters that it includes, the aim of this section is to present the needed datasets. As the studied country in this thesis is France, the presented datasets are those from France.

The two required datasets are :

1. **Hourly load data** : The model needs hourly data of electricity production. This dataset is taken from [6] which gives the actual electricity production per production type (nuclear, solar, fossil gas, biomass, ...).
2. **Hourly temperature data** : The model needs hourly data of temperature in the area studied. This dataset is taken from European Commission [14]. The chosen place to extract the data is Clermont-Ferrand[12] since it is approximately at the center of France.

Once these datasets have been extracted, they unfortunately cannot be used directly. Indeed, because of the time change in March there is one hour to remove because there is no available data (since the hour goes directly from 01:59 to 3:00, the data for 02:00 does not exist). Moreover, in the temperature dataset there is only one data for the time change in October (in the load dataset there are two data because the hour goes from 02:59 to 02:00) so one hour has to be removed in the load dataset : the choice is abritrarily made on the "second" 02:00-03:00 of the October time change.

These hours are not the only ones to remove, indeed sometimes there is no sufficient load data to use the model or there is an exceptionnally low electricity load which can

---

[12] This choice is discussed in section 6.

deviate the regression. This is shown on figure 3.7 : there are several points much smaller than the curve and others very near to zero or even equal to 0. All these anomalies have to be removed in order to have a good computation when resolving the regression. All the figures presented in section 3.2 were figures when the cleaning had already been done and the figure 3.1 is the figure 3.7 once the anomalies have been removed.

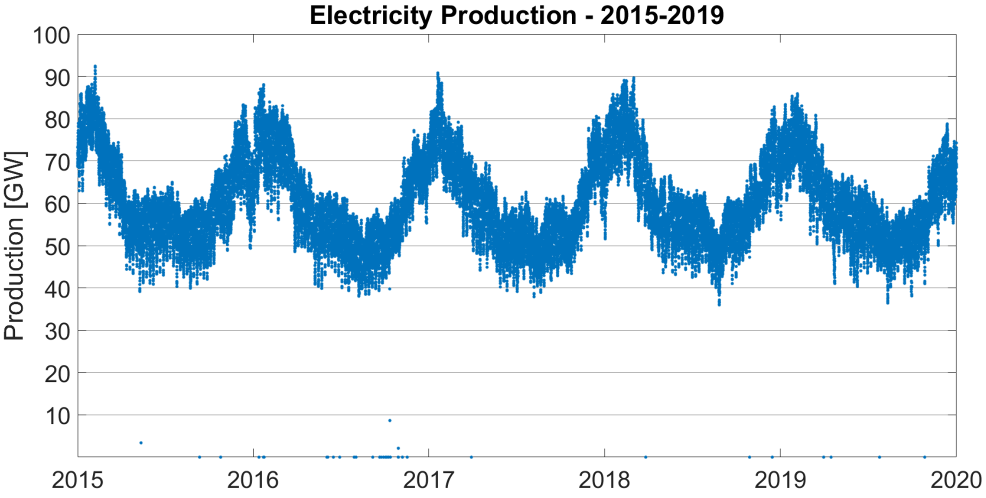


Figure 3.7: Electricity production in France from 2015 to 2019.

3.5 Performance Evaluation

Now that the model and the datasets have been presented in sections 3.2, 3.3 and 3.4, one can wonder how to evaluate the accuracy of this model. Indeed one of the main goals of a forecasting model is to be accurate enough, otherwise it is not of any interest. For the example, let's use the general form of a prediction for a set of observation Y_i : a set of predictors X_i. The prediction is given by :

$$\hat{Y}_i \; = \; \beta_0 \; X_{i,0} \; + \; \beta_1 \; X_{i,1} \; + \; ... \; + \; \beta_n \; X_{i,n} \tag{3.5}$$

There exist several different methods to evaluate the error of a prediction of the form of equation 3.5 and the most common are : Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE). These are given below :

$$MAE \; = \; \frac{1}{n} \sum_{i=1}^{n} \mid \hat{Y}_i \; - \; Y_i \mid \tag{3.6}$$

$$MAPE \; = \; \frac{100}{n} \sum_{i=1}^{n} \frac{\mid \hat{Y}_i \; - \; Y_i \mid}{Y_i} \tag{3.7}$$

$$MSE \; = \; \frac{1}{n} \sum_{i=1}^{n} (\; \hat{Y}_i \; - \; Y_i \;)^2 \tag{3.8}$$

In the equations 3.6, 3.7 and 3.8, the most easy to interpret is the MAPE because it compares the error of prediction with the observation. This allows to give an error expressed in percentage, which is more relevant than an error which is scaled to the observations or a squared error (especially in this case because even a small percentage error leads to very high absolute or squared error because of the high values of electricity load).

The chosen evaluation measure which will be used in the section 3.6 is thus the MAPE.

3.6 Validation Method

Now that the model has been completely presented in the previous sections, it is time to proceed to the validation to show that it works. The aim of this section is to present the used method to validate the performances of the model.

3.6.1 Duration of Calibration

First of all one can wonder how many data are needed in order to have an accurate enough prediction. Since the load dataset is available from 2015 and since the main goal of this thesis is to analyze the impact of COVID-19, the maximum duration to calibrate the coefficients of the model is 5 years : from 2015 to 2019. According to [15], only 2 years of data are required to have results.

Let's analyze the evolution of the performance of the model in function of the number of calibration years. In order to do this, the model is used to forecast successively the years 2016, 2017, 2018 and 2019, using all the available years before : for example, for the forecasting of 2016, only 2015 is used but for the forecasting of 2018, the years 2015, 2016 and 2017 are used.

The table 3.2 resumes the MAPEs for the different tests. These MAPEs are expressed in percentage of the actual production and the table is constructed as follows : the first column gives the calibration years, then the rows give the MAPEs obtained for the different years of the forecasting and the most important MAPE is of course the one of the forecasted year. This forecasted year is the last one of each row and is written in bold. Then the last column gives the overall MAPE, i.e. the MAPE computed on all the calibration years and not one at the time as in the other columns.

	2015	2016	2017	2018	2019	Overall
2015	3.04	**12.45**	/	/	/	3.04
2015-2016	3.44	3.62	**5.33**	/	/	3.53
2015-2017	3.65	3.84	3.82	**6.24**	/	3.77
2015-2018	3.98	4.02	4.04	3.76	**3.84**	3.95

Table 3.2: The MAPEs (expressed in %) for the successive forecasted years : 2016, 2017, 2018 and 2019. The left column gives the calibration years used to forecast the next and the rows give the computed MAPEs for the calibration years and for the forecasted year (the row "2015-2016" means that 2017 is the forecasted year).

There are two main remarks to do about the table 3.2 :

1. The overall MAPE is increasing with the number of calibration years. This is logical because the number of coefficients of the model is the same so with more data they cannot be as accurate as if they are computed for only one year. However this increase is not large : less than 1% of increase between the use of one year and the use of 4 years. Moreover the increase seems to be getting smaller and smaller : first there is an increase of nearly 0.5%, then nearly 0.24% and finally less than 0.2%. So we could be expecting this increase to stabilize at a moment.
2. Except a litte increase between 2017 and 2018, the MAPE of the forecasted year decreases with the number of calibration years. The statement of [15] which says that the model requires at least two years of data seems correct : with only one year the error is too large to have a good prediction model but as soon as there are two years used, the error gets a reasonable value and gets even smaller when all available years are used.

These two observations are illustred on figures 3.8 and 3.9 which respectively show the overall relative error during the 4 durations of calibration and the relative error per forecasted year. The figure 3.8 shows that the overall error is effectively increasing when the number of calibration years is increasing : there are much more peaks above 20% and most of the curve is above 10% of error for 4 years of calibration whereas for on year most of the curve is below 10% of error. The figure 3.9 shows that the relative error per year is effectively decreasing with the increasing number of calibration years : most of the curve is under 10% of error for 2019 whereas most of the 2016 curve is between 10% and 20% of error.

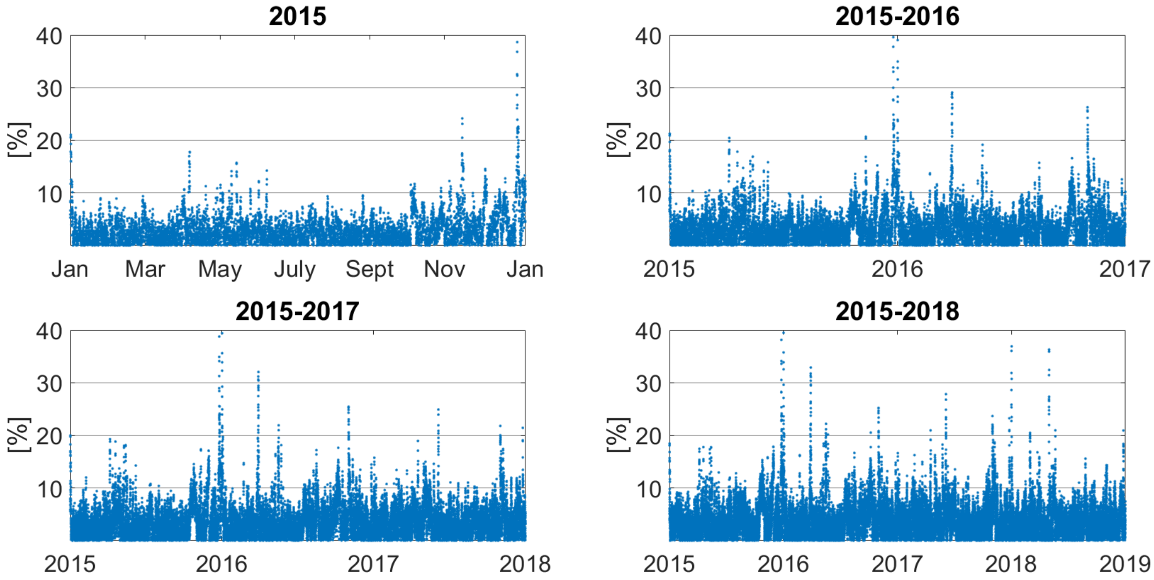


Figure 3.8: Overall relative error for the different durations of calibration. The x-axis is just there to show the evolution of the error during the calibration duration and the y-axis represents the error expressed in %.

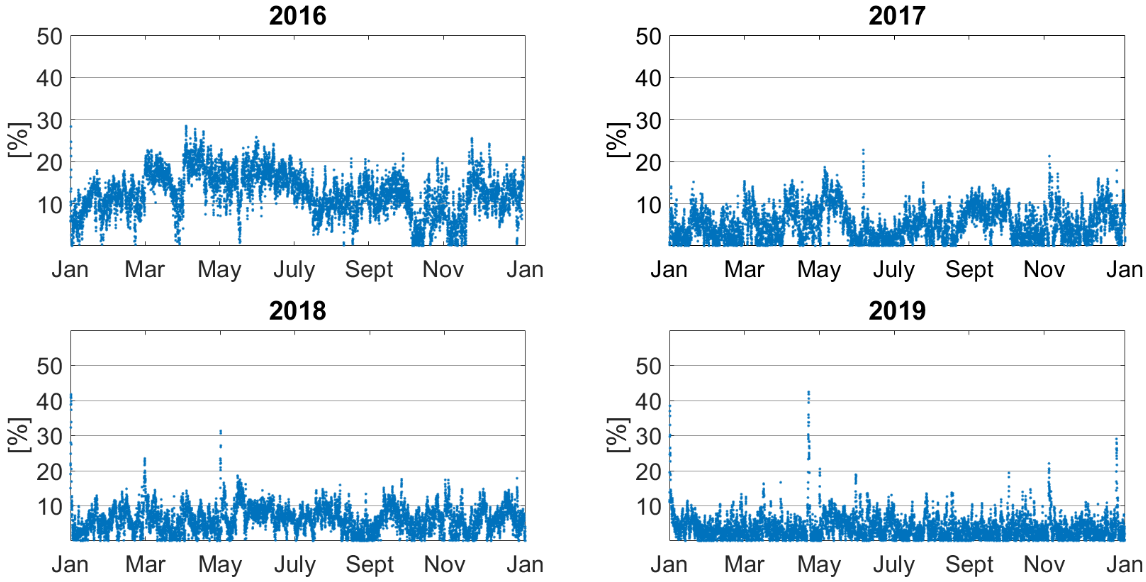
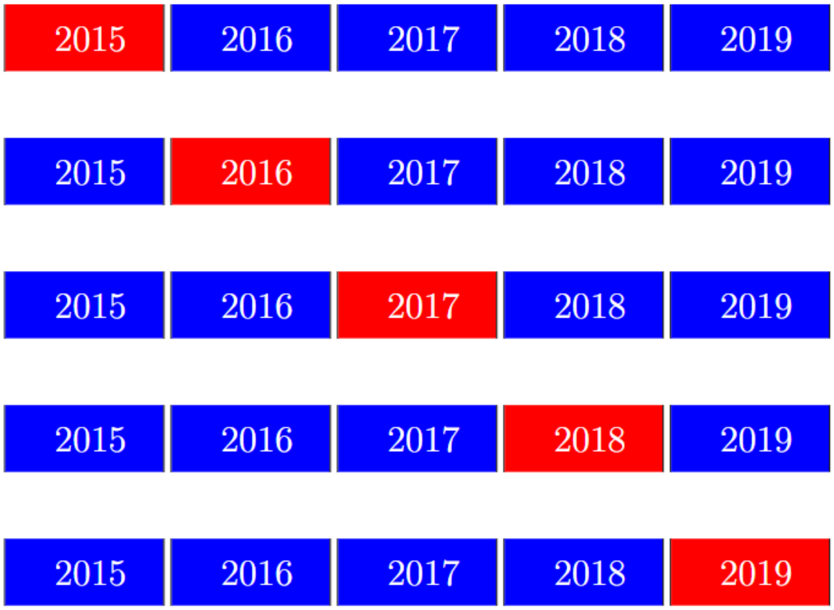


Figure 3.9: Relative error per year. The x-axis is just there to show the evolution of the error during the forecasted year and the y-axis represents the error expressed in %.

So, although the overall error is increasing when the duration of calibration is increasing, the error of the forecasted year is decreasing at the same time. Since the increase of the overall error is not so important and since the aim of the model is to make a prediction as accuracy as possible, the choice is made to use all the available years.

3.6.2 Cross - Validation

Now that the duration of calibration has been studied and chosen, a cross-validation can be done in order to show that the model works not only when the years are used in chronological order. The idea [16] behind the cross-validation is to test the model on each year of the dataset : 2015 is forecasted with years 2016 to 2019 as calibration years, 2016 is forecasted with years 2015 and 2017 to 2019 as calibration years. The principle is resumed on figure 3.10 with the calibration years in blue and the forecasted year in red.



2015	2016	2017	2018	2019
2015	2016	2017	2018	2019
2015	2016	2017	2018	2019
2015	2016	2017	2018	2019
2015	2016	2017	2018	2019

Figure 3.10: Principle of cross-validation.

The results of this cross-validation are given in the table 3.3.

Training period	**Forecasted year**	**MAPE [%]**
2016-2019	2015	5.48
2015 and 2017-2019	2016	4.69
2015-2016 and 2018-2019	2017	4.49
2015-2017 and 2019	2018	4.27
2015-2018	2019	3.84

Table 3.3: Results of the cross-validation.

The main remark to make on the table 3.3 is that the MAPE is decreasing when the calibration years are mostly anterior years : there is a difference of about 1.7% between the MAPE of the forecast of 2019 and the MAPE of the forecast of 2015. This can be explained by the definition of the trend and the assigning of an increasing natural number to the whole dataset, beginning with the first hour of 2015.

The error of the previous cross-validation is not too large and so the model can be validated but the definition of the trend can also be modified to proceed to another cross-validation. Indeed, in order to consider the forecasted year as a real forecasted year, this year can be placed after all the others in chronological order : for the forecast of 2016, the chronological order for the trend is 2015, 2017, 2018, 2019 and then 2016. By doing this, the value assigned to $Trend$ is an increasing natural number, starting from the first hour of 2015, then continuing directly in 2017 in place of 2016 and so on until the end of 2019 when it continues for the year 2016 as it was the year after 2019. Since an illustration is often easier to understand a concept, this idea is represented on figure 3.11.

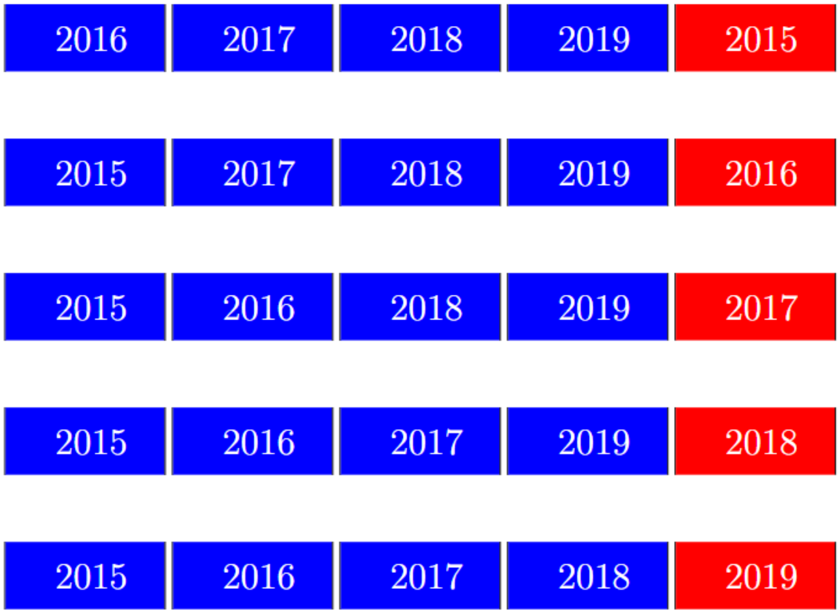


Figure 3.11: Principle of cross-validation with a modified trend.

The results obtained with this modified trend are peesented in the table 3.4.

Training period	**Forecasted year**	**MAPE [%]**
2016-2019	2015	4.31
2015 and 2017-2019	2016	4.29
2015-2016 and 2018-2019	2017	4.32
2015-2017 and 2019	2018	4.83
2015-2018	2019	3.84

Table 3.4: Results of the cross-validation with the modified trend.

The obtained MAPEs are smaller than the ones obtained without modifying the trend (except for the forecast of 2018) so it shows that the model works even to forecast a previous year. Thus the model is well validated and once again, the best prediction is for 2019 where the calibration years have been used in the good order, which is a good point because it is the way it will be used in section 4.

The aim of this whole section was to present a general forecasting model, explain how to use its parameters, determine how to evaluate its performances and finally validate it by evaluating its performances. Now that all that is done, the question of the modelization of the COVID-19 pandemic can be studied and this is what is done in section 4.

4 COVID-19 Shock Modeling

4.1 Introduction

Now that the general forecasting model has been presented, the next step (which is the aim of this section) is to determine how one can represent the impact of the COVID-19 pandemic. After a little check on the Internet, it is an evidence that it can be represented as an economic shock. Indeed, here is a definition of an economic shock given in [17] : "An economic shock refers to any change to fundamental macroeconomic variables or relationships that has a substantial effect on macroeconomic outcomes and measures of economic performance, such as unemployment, consumption, and inflation. Shocks are often unpredictable and are usually the result of events thought to be beyond the scope of normal economic transactions.".

With this definition, it is clear that the COVID-19 pandemic resulted in an economic shock and one can wonder how to modelize and study such an economic shock. Since this thesis focus on the electricity production, the question is thus to modelize a shock in terms of production.

As already explained in section 2.1, COVID-19 had as a consequence a period of low electricity consumption and so low production. According to [18], such a period of low production due to a shock can be represented as a negative impulse in a dynamic system which evolves in closed loop, as represented on figure 4.1.

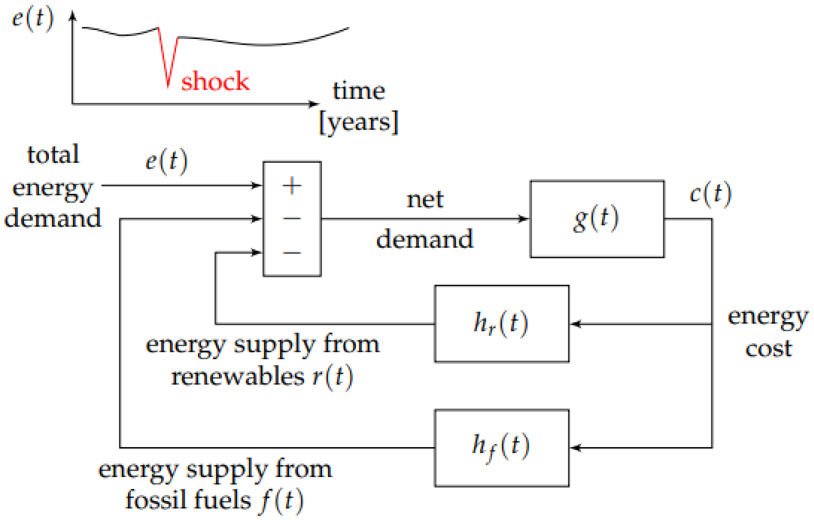


Figure 4. An illustrative feedback system describing the effects of "economic shocks": periods of low consumption are modeled as negative impulse signals, which cause reactions and counter-reactions that evolve in a closed feedback loop.

Figure 4.1: Representation of a shock on energy demand, from [18].

In figure 4.1, there are 3 black boxes : $g(t)$, $h_r(t)$ and $h_f(t)$ which link the net energy demand to the energy cost for $g(t)$ and the energy cost to respectively the energy supply from renewables and from fossil fuels for $h_r(t)$ and $h_f(t)$. Since it seems complicated at first sight to determine these functions, there is a need of a more specific research on the subject of dynamics of energy.

Two articles were found about that subject : [19] which is about the system of energy demand and supply in Canada and [20] which is about the development of a dynamic energy system in Australia. The problem with these articles is that they take into account some parameters such as, for example, the Research and Developments Investment or the Capital Assets Investment, the Electricity Production Cost or the Electricity Capacity. Although some of these parameters can be determined in this work (for example the electriciy capacity), some others are out of reach in the scope of this work. For illustration, the system studied in [20] is represented on figure 4.2.

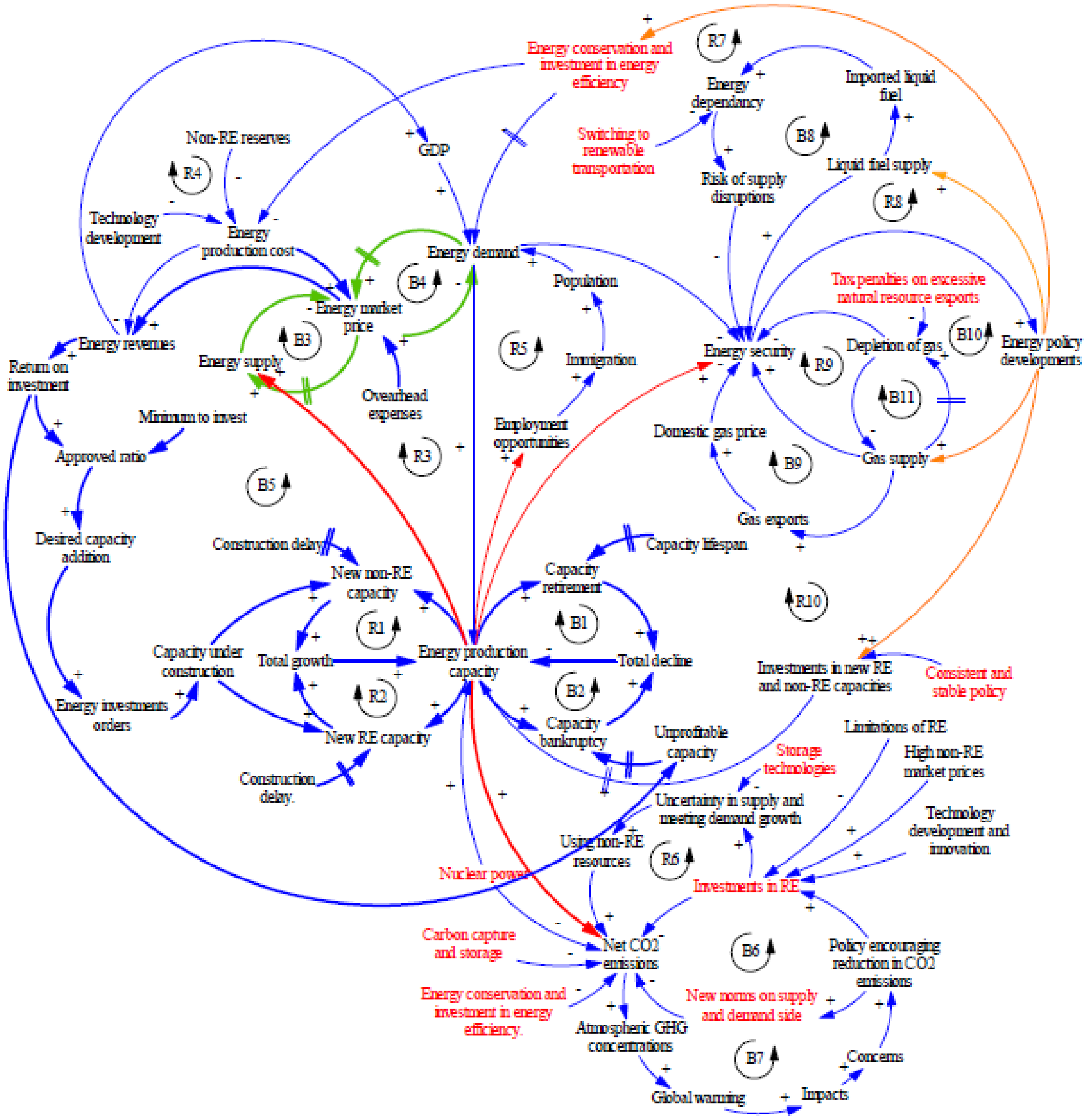


Figure 2. Feedback loops from Laimon et al. [1] causal loop diagram (CLD) replicated in the stock-flow model (indicated in bold). The CLD contains 21 feedback loops including ten reinforcing loops (R1 to R10) and eleven balancing loops (B1 to B11). This CLD highlights the main components of the energy sector linked to Australia's energy policy including energy resources (loops R1 and R2); energy production, supply and demand (loops R3, B3 and B4); energy economics (loops B5 and R4); energy emissions and energy emissions policies (loops R6, B6 and B7); and energy policy developments (loops R7, B8, R8, B9, R9, B10, R10 and B11). Parameters in red are missing or poorly performing in the Australian context.

Figure 4.2: Representation of a dynamic energy system in development, from [20].

The system represented on figure 4.2 is very complex and is thus beyond the scope of this work, such as the one studied in [19] but the idea behind theses articles is kept : the modeling of an electricity system as a dynamic system[13] to include the COVID-19 shock.

Now that the idea of the representation of a shock in an energy system is determined, the question of the precise modeling can be studied. In a dynamic model, there is often a state-space representation of the parameters to be studied. Indeed this representation reflects well the dynamics of a system with the state equation and the observation equation. A method that is widely used in the field of load forecasting (and forecasting in general) is the autoregressive model (AR model) and it can be used in state-space representation. All that is developed in section 4.2.

---

[13] First seen in [18].

4.2 State-Space Model and AR Process

This section presents the state-space and autoregressive process which will be used in this thesis. First a short literature review verifies that the autoregressive model is effectively used in the field of forecasting. Then the basics of state-space and autoregressive models are reminded and the way they can be linked to each other is presented. Finally the application to the case of this work is explained.

4.2.1 Literature Review

First a research in the literature was made to see if the autoregressive model is effectively used in the field of forecasting. Several articles were quickly found : for example Nidal S. Kamel, a professor at University Technology PETRONAS in Malaysia [21] has made several works on the subject (for example [22],[23] and [24]) with Zuhairi Baharudin, a PhD student (at the time of these works) in the same university.

All these works are based on the autoregressive model, with some modifications, but the fact is that they are all similar in a point : they give accurate results. Indeed in each work, the MAPE obtained in the experiment made is less than 5% for the basic autoregressive model and even smaller for the modified models (smaller than 2% in each work). Since Nidal Kamel has made many works and is cited more than 2000 times ([21]), one can guess that his work is pertinent and the conclusion is that the autoregressive model effectively works in the field of forecasting.

4.2.2 Theoretical Background

The general form of a state space model[14] [25] is given by equation 4.1.

$$\begin{cases} x[k+1] \;=\; A[k]\; x[k] \;+ B[k]\; u[k] \\[2ex] y[k] \;=\; C[k]\; x[k] \;+ D[k]\; u[k] \end{cases} \tag{4.1}$$

In equation 4.1, the parameters are [25] :

- $x[k]$: the state vector.
- $y[k]$: the output vector.
- $u[k]$: the input vector.
- $A[k]$: the state matrix.
- $B[k]$: the input matrix.
- $C[k]$: the output matrix.
- $D[k]$: the feedthrough matrix.

The more general form of an AR model [26] is the ARMAX (autoregressive moving average with exogenous inputs) model, given by equation 4.2.

$$\begin{aligned} Y(k) \;+\; a_1\, Y(k-1) \;+\; ... \;+\; a_n\, Y(k-n) \;=\; b_0\, U(k) \;+\; ... \;+\; b_m\, U(k-m) \\ +\; E(k) \;+\; c_1\, E(k-1) \;+\; ... \;+\; c_p\, E(k-p) \end{aligned} \tag{4.2}$$

---

[14] Written in discrete form, because this is the form needed in this work.

In equation 4.2, the parameters are :

- $Y(k)$: the autoregressive parameter.
- $B(k)$: the exogenous input.
- $E(k)$: a white noise representing the moving average part of the process (the "random" part).
- a_i : the coefficients of the autoregressive part.
- b_i : the coefficients of the exogenous inputs (the "deterministic" part).
- c_i : the coefficients of the moving average part.

One can notice that equations 4.1 and 4.2 can be linked togheter. This is what is done for example in [27] : the equation 4.2 can be rewritten in matricial form, and if the process is considered to have no moving average part (i.e. $c_i = 0$) and that $U(k-i)$ is only included for $i = 0$:

$$\bar{Y}(k) \;=\; A\,\bar{Y}(k-1) \;+\; B\,\bar{U}(k) \tag{4.3}$$

In equation 4.3, $\bar{Y}(k)$, A and B are given by :

$$\bar{Y}(k) = \begin{bmatrix} Y(k) \\ Y(k-1) \\ ... \\ Y(k-n+1) \end{bmatrix} \quad , \quad A = \begin{bmatrix} -a_1 & -a_2 & ... & -a_{n-1} & -a_n \\ 1 & 0 & ... & 0 & 0 \\ 0 & 1 & ... & 0 & 0 \\ 0 & 0 & ... & 0 & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & ... & 1 & 0 \end{bmatrix} \quad , \quad B = \begin{bmatrix} b_0 \\ b_1 \\ ... \\ b_m \end{bmatrix} \tag{4.4}$$

With the equation 4.3 written in matricial form, the link with the state space representation 4.1 becomes evident, the only need is to add the observation equation in order to have a state space representation from the ARX (autoregressive with exogenous input) model.

The next sections explains how to use these models in the case studied.

4.2.3 Application to the Case of this Work

Now that the basics of state space representation and autoregressive model are reminded, the next step is to present how to use it in the case of this thesis.

First of all the ARMAX process is considered, as presented in equation 4.2. Let's rewrite this equation here for more clarity :

$$\begin{aligned} Y(k) \;+\; a_1\, Y(k-1) \;+\; ... \;+\; a_n\, Y(k-n) \;=\; b_0\, U(k) \;+\; ... \;+\; b_m\, U(k-m) \\ +\; E(k) \;+\; c_1\, E(k-1) \;+\; ... \;+\; c_p\, E(k-p) \end{aligned}$$

In this work the autoregressive part of the process $Y(k)$ is simply the state vector of the different predictions at the different times of the study.

The exogenous input $U(k)$ is the prediction made with the general model (presented in section 3) since the objective is to be able to predict the impact of a shock in advance. Only $U(k)$ is considered, $U(k-i)$ are not included because the input signal is considered to be causal.

The moving average part $E(k)$ is not taken into account because it can be represented as the ”random” part of the process, whereas the exogenous input is the deterministic part of the process. In fact, in this work, the exogenous input can be decomposed into a deterministic and a random part. Indeed, as explained in section 3.2, the model is based on calendar variables, temperature and interaction effects between them. The calendar variables can be assimilated to the deterministic part of the process, the expected production in function of the time of the year. On the other hand, the temperature can be assimilated to the random part of the process because, although there is generally a trend all over the year, temperature can sometimes be very random.So the exogenous input $U(k)$ is in fact decomposed into two components : $U_c(k)$[15] and $U_t(k)$[16], which is the reason why the moving average part is not taken into account.

Now that the parameters of the ARX model that will be used have been presented, they can be included into the state space representation as it is done in section 4.2.2. In the case of this work, the equation 4.1 becomes :

$$\begin{cases} x[k+1] \ = \ A[k]\ x[k] \ + B[k]\ u[k] \\ y[k] \ = \ C[k]\ x[k] \end{cases} \tag{4.5}$$

In equation 4.5, $x[k]$[17] is the state vector at time k, $u[k]$[18] is the input vector at time k :

$$x[k] = \begin{bmatrix} x[k] \\ x[k-1] \\ \dots \\ x[k-n+1] \end{bmatrix} \quad , \quad u[k] = \begin{bmatrix} u_c[k] \\ u_t[k] \end{bmatrix} \tag{4.6}$$

Concerning the matrix A, B and C, A is exactly the same matrix as in equation 4.3, B must be modified since $u[k]$ must only impact the last value of $x[k]$ and not the previous ones and C can be chosen arbitrarily to have an outpout $y[k]$ which is the last value of $x[k]$ computed :

---

[15] c stands for calendar.

[16] t stands for temperature.

[17] $Dim(x) = n \times 1$

[18] $Dim(u) = 2 \times 1$

$$A = \begin{bmatrix} -a_1 & -a_2 & ... & -a_{n-1} & -a_n \\ 1 & 0 & ... & 0 & 0 \\ 0 & 1 & ... & 0 & 0 \\ 0 & 0 & ... & 0 & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & ... & 1 & 0 \end{bmatrix} \quad , \quad B = \begin{bmatrix} b_c & b_t \\ 0 & 0 \\ ... & ... \\ 0 & 0 \end{bmatrix} \quad , \quad C = \begin{bmatrix} 1 & 0 & ... & 0 \end{bmatrix} \tag{4.7}$$

In equation 4.7, all the a_i, b_c and b_t coefficients can be computed with least squares for example, using a training dataset as given in section 3.4.

4.3 Modelization and Validation Method

Now that the model has been completely presented in the previous sections, it is time to proceed to the validation to show that it works.

4.3.1 Order of the Model

First there is the question of the number of the coefficients needed to have an accurate enough model : this is the order of an AR model (i.e. the a_i coefficients of matrix A in equation 4.7). One can have the intuition that the more coefficients there are, the more accurate the model will be, but it's not necessarily true, there can be an optimal order of the model to have the best accuracy. Moreover, if the order is very high, the computation time would also increase, which is not wanted. Thus this section evaluates the optimal order of the model, also taking into account the computation time. This evaluation is done with the dataset presented in section 3.4, using the year 2018 as training year and the year 2019 as validation year.

The computation time of the method used is presented on figure 4.3 for a large range of order, which goes from 1 to 336 (whichs means two whole weeks of data to forecast the next hour). One can notice a global increase of the computation time starting about an order of 96 and then a slowly increasing trend, followed by a peak between an order of 168 and 192 and finally a drop after an order of about 300. A logical conclusion if the computation time is important would thus be to choose an order smaller than 96 or higher than 300 since these are the two regions of the graph where the computation time is the lowest.

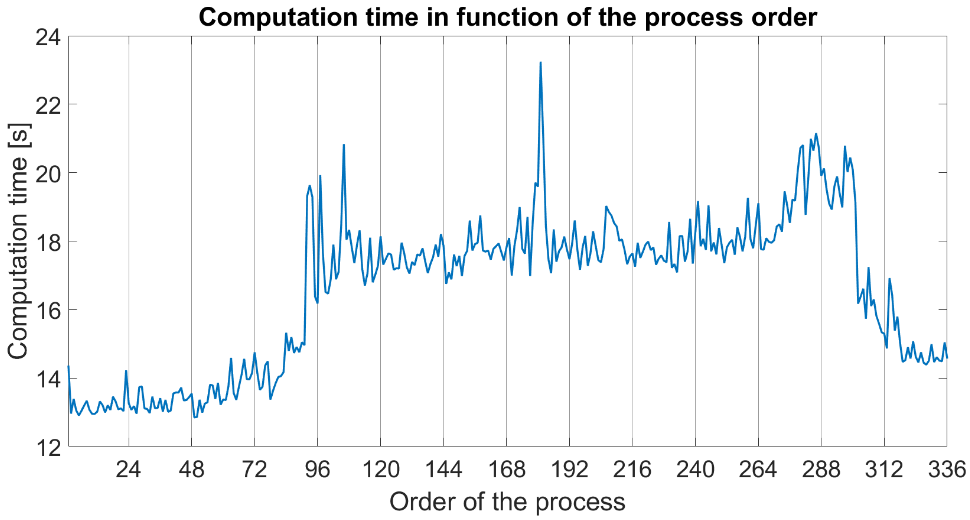


Figure 4.3: Evolution of the computation time of a whole year prediction with the process order.

However, the differences between the different times are not so relevant : the shortest time is about 13 s and the longest is about 23 s. Since, in this case, the program should not be executed into a loop, there is no real problem in having a computation time of about 20 s. Moreover, the computation time of a program can strongly change from one execution to another, thus the computation time is here not really important[19] and the accuracy is the determinant parameter to take into account when choosing the process order. This accuracy is determined taking the MAPE of the forecasted year (see section 3.5). The evolution of the MAPE with the order of the process is presented on figure 4.4.

One can notice that the MAPE of the model is not monotonous and does not decrease continually. Indeed, although there is a global decreasing trend with the increasing process order, there are some peaks but two of them are major : whereas the MAPE seems to decrease strongly at the first order changes, there is a first peak which nearly

---

[19] However it would be important if this model was used in a big loop, for further analysis for example.

brings the curve back to its beginning value. The second and maybe most important peak is at the center of the graph : whereas one would expect the MAPE to decrease when the process order increases, when this order reaches a value of about 150 and to 170, the MAPE increases to reach the highest value of the curve.

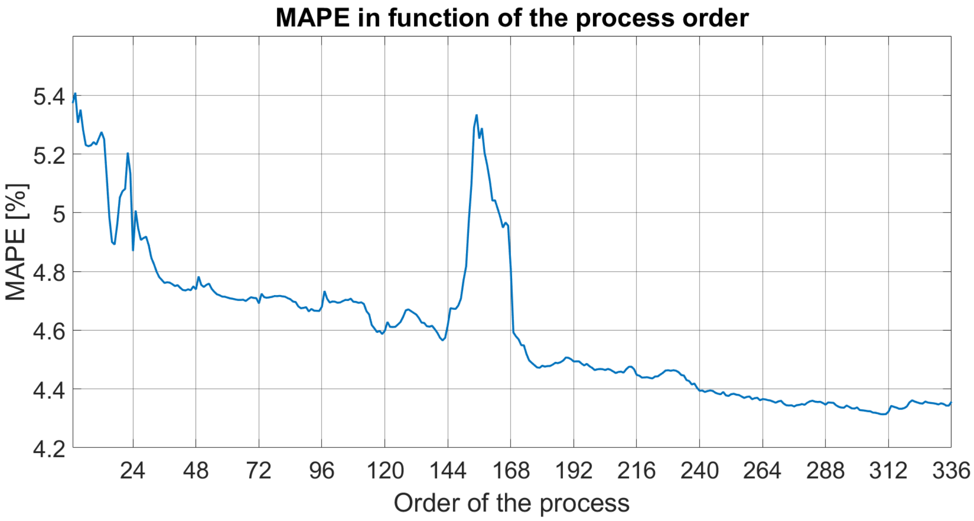


Figure 4.4: Evolution of the MAPE of a whole year prediction with the process order.

Thus, as mentioned in the beginning of this section, a process of high order does not necessarily means a good[20] accuracy. The optimal value is at the order 309 for a MAPE of 4.314%. Since it can be more relevant to talk in terms of days than in terms of order[21], the table 4.1 resumes the obtained MAPE for 1 day of data to two weeks of data.

---

[20] Rather "the best" than "good", about 5% of error is still acceptable.

[21] A process of order 24 means a day of data to forecast the next hour, order 48 means two days,...

Process order	**MAPE [%]**	**Process order**	**MAPE [%]**
24	4.87	192	4.49
48	4.74	216	4.45
72	4.69	240	4.39
96	4.68	264	4.37
120	4.60	288	4.35
144	4.62	312	4.32
168	4.81	336	4.36

Table 4.1: Different MAPEs of the model with their respective order (which represent 1 to 14 days of data used to forecast the next hour).

With the table 4.1, if the process order must be chosen in order to be assimilated to a day, two days, etc. of data, the optimal order is 312 (thus 13 days of data). However, bewteen an order of 24 and 168 and between an order of 192 to 336, all the MAPEs are close to each other with variations of the order of only 0.1% so the order could be chosen arbitrarily or with another parameter.

Indeed, another parameter had to be included after first choosing an order of 312 and this paragraph explains the reason why. The aim of this model is to be able to include a shock easily[22] and the problem of choosing an order too high does not allow that with good accuracy. Indeed, after making some tests, an order smaller than 168 has a better accuracy when including a shock. This can be explained by looking at the coefficients of the model. The figure 4.5 shows these coefficients with respect to their indices and the figure 4.6 shows the same coefficients but without the first two, in order to have a better scale for the others.

On the figure 4.6, one can see that most coefficients are included between -0.5 and 0.5

---

[22] That is developed in section 4.3.2.

with some exceptions. This is the reason why a model with many coefficients is less appropriate to include a shock than a model with less coefficients. Indeed, the value of the b coefficients, i.e. the coefficients of the input signal are about 0.04. Thus, if more and more coefficients are added in the model, the impact of the input signal decreases at each new coefficient. However, as it will be explained in the next section, the input signal must keep an important impact on the model to modelize a shock.

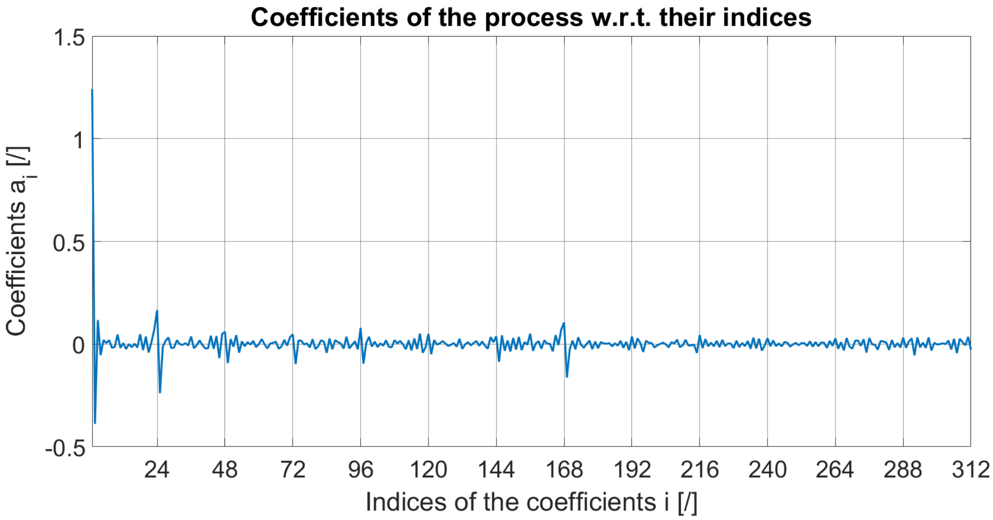


Figure 4.5: Coefficients of the process w.r.t. their indices.

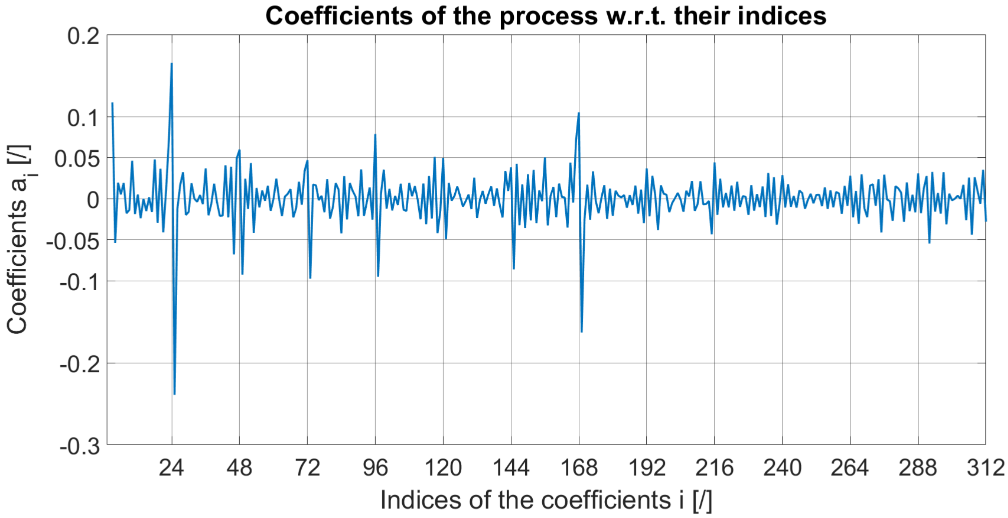


Figure 4.6: Coefficients of the process w.r.t. their indices without the first two coefficients.

With this fact and the MAPEs resumed in table 4.1, since it is the optimal order in terms of accuracy, the chosen order is 120, so 5 five days of data to forecast the next hour. Thus the computation time is not taken into account in this work but it could be if this model was used in another study.

To conclude this section, the presented MAPEs do not exceed a value of 5%, except for some peaks, so the model can be validated and the next step can be reached : including a shock as the COVID-19 one.

4.3.2 Including the COVID-19 Shock

Now that the order of the model is determined and that its accuracy has been shown, the question of the inclusion of the shock must be studied. Indeed, since the MAPEs presented in the section 4.3.1 are greater that the ones presented in section 3.6, one can wonder why use the model presented in section 3 only as an input signal in the model presented in this section and not use it directly. There are two reasons to this way of using it.

First, in the general model, including a shock is complex because of its definition. Since it only includes calendar and temperature effects, including a shock is not easy. On the other hand, with the AR model expressed in state space, a shock can be modeled at any time as a negative impulse or a input signal modified to correspond to this shock. Indeed, if the event that causes the shock can be translated in clear modifications of the electricity production[23], the input signal can be modified easily to follow the evolution of this shock.

Then, if the analysis[24] must be done on the long term, there is a need of an input signal accurate enough and that is when the general model is also useful. Indeed, since its accuracy has been proven, it can be used as a good input signal for a complete year of forecasting. With this input signal and the AR model expressed in state space, once it is trained, only one week of electricity data are necessary to study one year of forecast[25]. This fact and the fact that a shock can be included in the model if this shock can be represented easily in terms of electricity production modifications are the two reasons

---

[23] For example an arbitrary drop of 5% of the electricity production due to political decisions (this is obviously hypothetical but it's to present a case easy to modelize).

[24] A warning must be done here : the analysis is on the long term but not the forecast, since the input signal needs temperature data which cannot be accurate enough when forecasted too far ahead.

[25] Considering that the data used to train the AR model and the input signal are not counted in the "one week".

of the development of this state space AR model using a general model as input signal.

4.3.3 Application to the Year 2020

Now that the modelization of a shock has been explained, the model can be applied to the case of this thesis : the COVID-19 pandemic during the year 2020. In France, the first lockdown began on the 17th March[28] and so this is the date acknowledged as the first shock. The end of the lockdown is on the 22th June and so it could in theory be acknowledged as a return to normal production.

To modelize this lockdown, a first approach is to simply reduce the input signal by taking a precentage of it. It is not rigorous but, since it is obvious that a lockdown results in a drop of electricity consumption (and so production), it should work as a first approach. After some tests, errors and corrections, taking an input value of 85% of the input signal leads to a good accuracy for the lockdown period, as shown on the figure 4.7 but one can see that taking the end of the lockdown as a return to previous conditions is wrong. It can be explained by the fact that most industries continued to apply working from home for their employees and so they need less electricity than a normal consumption. In fact it is rather around the 20th September that the input signal can recover 100% of its value to follow the curve. Indeed, when taking the modified input signal until the end of September rather than until the end of June, the curve is correctly followed, as shown on figure 4.8.

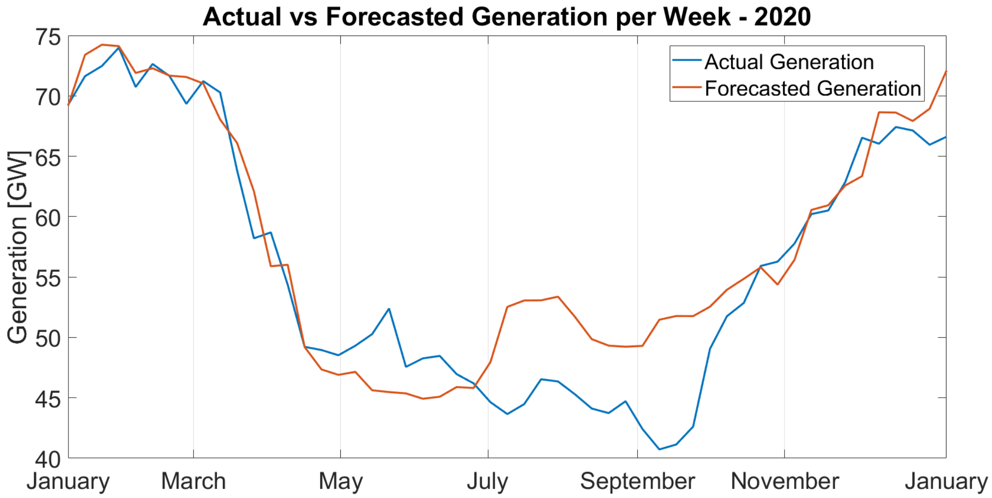


Figure 4.7: Actual and forecasted electricity generation per week in France in 2020, with the end of the shock in June.

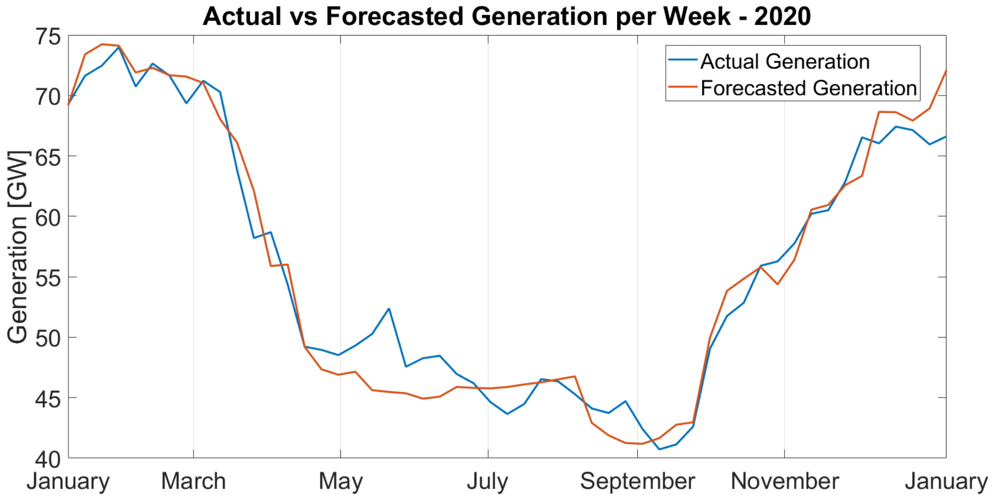


Figure 4.8: Actual and forecasted electricity generation per week in France in 2020, with the end of the Shock in September.

Just to compare the efficiency of the two tests between them and with the forecasting without including the shock, the figure 4.9 shows the actual and the forecasted production without including the shock in the model, and the table 4.2 presents the different MAPEs obtained.

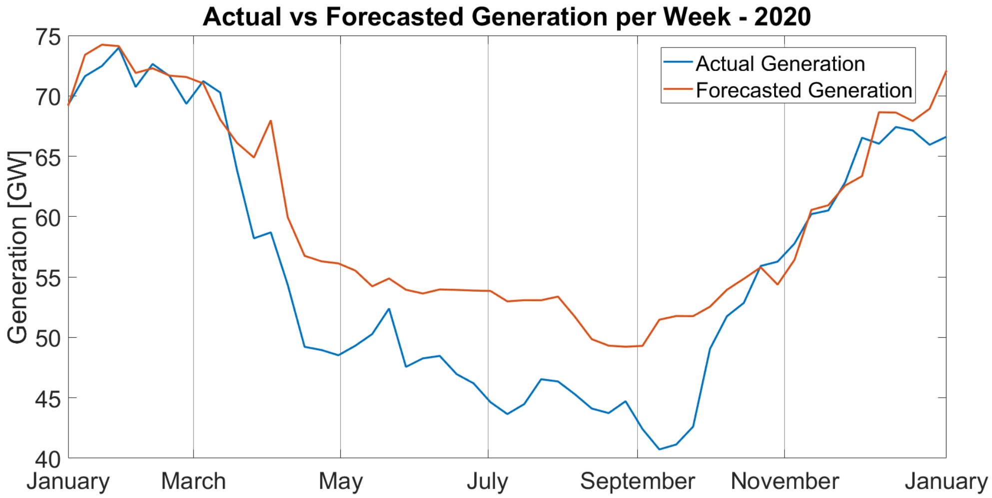


Figure 4.9: Actual and forecasted electricity generation per week in France in 2020.

Duration of the shock	**MAPE [%]**
/	10.21
Until June	8.99
Until September	6.37

Table 4.2: MAPE of the forecasting of 2020 with the way of modeling the shock.

Now that the model is shown to be able to include the COVID-19 shock, the section 4.4 presents some results obtained with this model.

4.4 Results

The results presented in this section are about the total production during the year 2020, because it can be more relevant to analyse the drop due to the pandemic with the total production of the year. In 2019, the total production[26] was 524.97 TWh. In 2020 the production was 488.98 TWh, which corresponds to a drop of 6.86% with respect to 2019. With the model including the shock, the total forecasted production in 2020 is 487.5 TWh, which corresponds to a drop of 7.14% w.r.t.[27] 2019 and a difference of 0.3% w.r.t. the actual production in 2020.

Now the model can be used to imagine an hypothetical lockdown at another period : since the shock due to this lockdown lasted about 6 months, one can imagine this shock happening the first or the last 6 months of 2020. For the first 6 months, the total production of the year would have been 482.11 TWh and for the last 6 months, it would have been 483.03 TWh. These predictions correspond to a drop of respectively 8.16% and 7.99% w.r.t. the production in 2019. A first conclusion to the fact that this drop is higher than the real drop is that the moment when the shock happens impacts the magnitude of the shock. Then, if the hypothetical drops are higher, it is because they happen in part during winter, so when the production is usually higher than during spring or summer.

The table 4.3 resumes these different drops. In this table, ”2020*” represents the year with an hypothetical shock starting the 1st January and lasting the same duration of the real shock and ”2020**” represents the year with an hypothetical shock ending the 31st December and lasting the same duration of the real shock (so starting around the beginning of June).

---

[26] All data are taken for France [6].

[27] With respect to

Year	**Total production [TWh]**	**Drop w.r.t. 2019 [%]**
2019	525	/
2020	489	6.9
2020*	482	8.2
2020**	483	8

Table 4.3: Drops of the production in 2020 w.r.t. 2019 for three differents scenarios.

Now the figure 4.10 presents the evolution of the drop of the production with the duration of the shock. This graph shows a linear evolution that goes up to more than 15% of drop for a whole year of shock period.

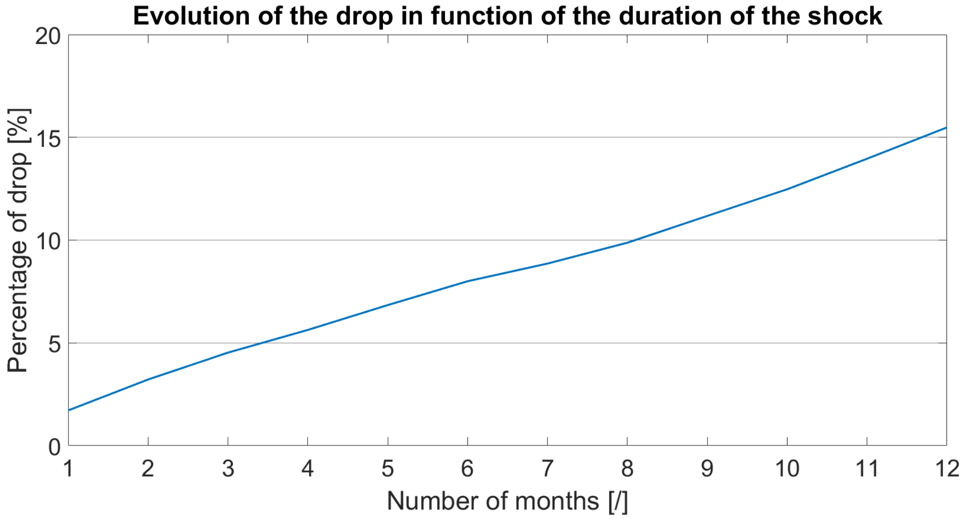


Figure 4.10: Evolution of the drop of the production with the duration of the shock.

Of course, the method used to have these results is not rigorous because it used some tests and corrections to have a correct model. This leads to the next question of this thesis : How to translate an event into a clear modification of the electricity production or how to determine the magnitude of a shock in function of its type ? This will be the subject of the section 5.

5 Discussion and Shock Analysis

To have a better understanding of the percentage of the initial value taken as input signal in section 4.3.3 and of the duration of this modified input, it is right to have a look first at the repartition of the electricity consumption between the different sectors and to the measures taken by the government. With these measures, one can determine how the different sectors were personally impacted. For the duration, one should look at the measures that continued even after the end of the lockdown (for example working from home).

The aim of this section is thus to make a short analysis of the modification of the electricity repartition in sectors, the measures taken by the government and the link between the two. The reader should pay attention to the fact that this analysis is a qualitative analysis and not a quantitative one. This is because some data or informations are out of reach in the case of this thesis, and some computations would need a specific research[28] in order to have precise results. Indeed, each time that the following paragraphs explain what could be done if an information was known, it is because this information could not be found in the scope of this work. Now that the precision about the type of analysis is done, the following paragraphs present this analysis.

First of all, economoy or energy are often separated in three distincts sectors[29] :

- The primary sector involves the extraction of raw materials : fishing, farming are in the primary sector for example.
- The secondary sector involves the transformation of the raw materials from primary sector into products : steel from mining into a kitchen tool, textile into clothes for example.

---

[28] Specific research which could be the subject of another work.

- The tertiary sector involves the supply of services to consumers : there are many categories of services[29] but several exampls can be banks, insurances, stores, hospitals, hotels/restaurants, education, etc.

Once this repartition is clearly made, it becomes easier to determine which sector has been more or less impacted by COVID-19 than the others. Indeed, the lockdown essentially impacted the tertiary sector : schools, all the hospitality industry, non essential stores were closed, leading to a drop of consumption in all these categories. On the other hand, all the health system (hospitals, pharmacies, particular physicians) was open and more required than the usual demand, leading to an increase of consumption.

Primary and secondary sectors has been less impacted than the tertiary : government defined "essential" and "non essential" jobs, this repartition was made to prevent non essential workers to go to work and thus to reduce the contacts between people. Thus, if the repartition between essential and non essential jobs in secondary[30] sector is known, the impact of the lockdown on these sectors can be directly determined. For example, if a company has only non essential workers who do not go to work during the lockdown, its electricity consumption can be taken as null or nearly null during this period, unless a certain constant consumption is required.

For the tertiary sector, the modifications are more complex to determine because of the categories that lead to a drop and the categories that lead to an increase of consumption. For this sector, some data are needed to make a good analysis : ideally data of the repartition of the consumption between all the categories of the sector. Indeed, if the usual electricity consumption in schools is known, then this consumption can be set to zero during the lockdown, and the same happens for all the services that were simply

---

[29] A more exhaustive list can be found on [30].

[30] Jobs in the primary sectors are mostly considered essential.

closed during the lockdown (hotels, restaurants, clothing stores for example). For the health system, the modification is more complex and could need a specific study just on it. For hospitals for example, a first approach can be to base on the number of simultaneously occupied beds during a usual period and to convert the consumption in terms of this number of patients. However, this implies also data of all the hospitals, or in the case of a forecast study, the prediction of the evolution of the number of patients during the pandemic.

Now that a global idea to determine the impact of a lockdown on the three main sectors has been presented, one can notice that an important part of the electricity consumption is missing : indeed, the residential part is not presented in the three sectors and it is a major part, as it is shown on figure 5.1.

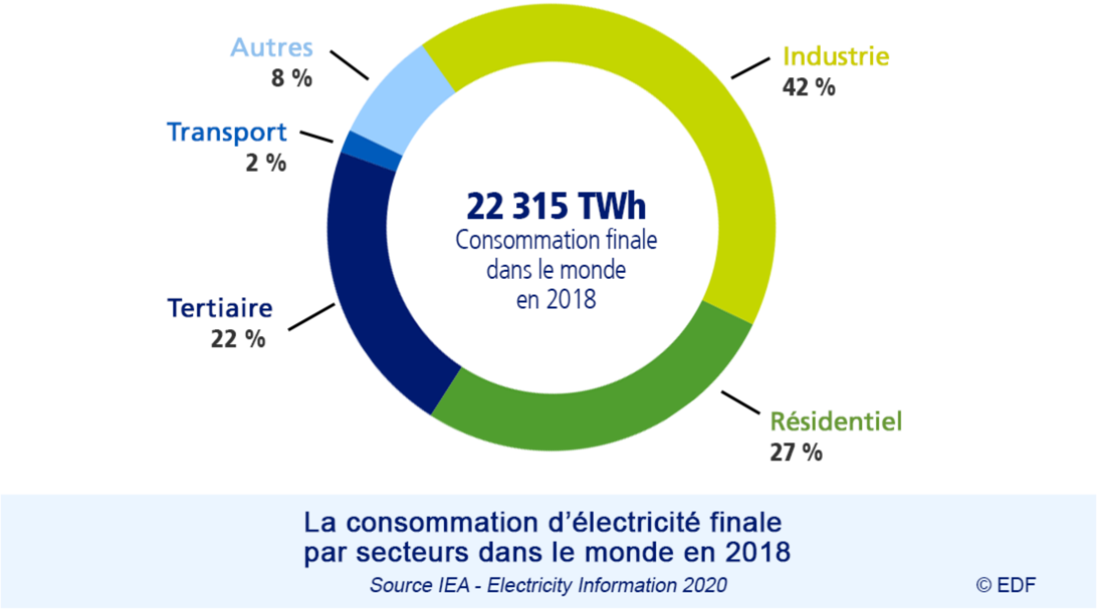


Figure 5.1: Repartition of the electricity consumption between different uses in the world in 2018.

This figure is taken from [31] and shows the repartition of the electricity consumption between different uses in the world in 2018. The three main parts are the tertiary sector which represents 22% of the consumption, industry (which can be considered as the mix of the primary and secondary sectors) which represents 42% and the residential consumption which represents 27%, i.e. about one third of the total consumption. Thus this figure shows that the residential is an important part of the consumption and that it must be taken into account in this qualitative analysis of the lockdown, especially because the residential consumption was expected to increase during this lockdown.

To evaluate the modifications due to the lockdown on the residential consumption, a knowledge about the way that electricity is globally used at home can be useful. A repartition of the use of electricity at home in France in 2019 is shown on figure 5.2, taken from [31].

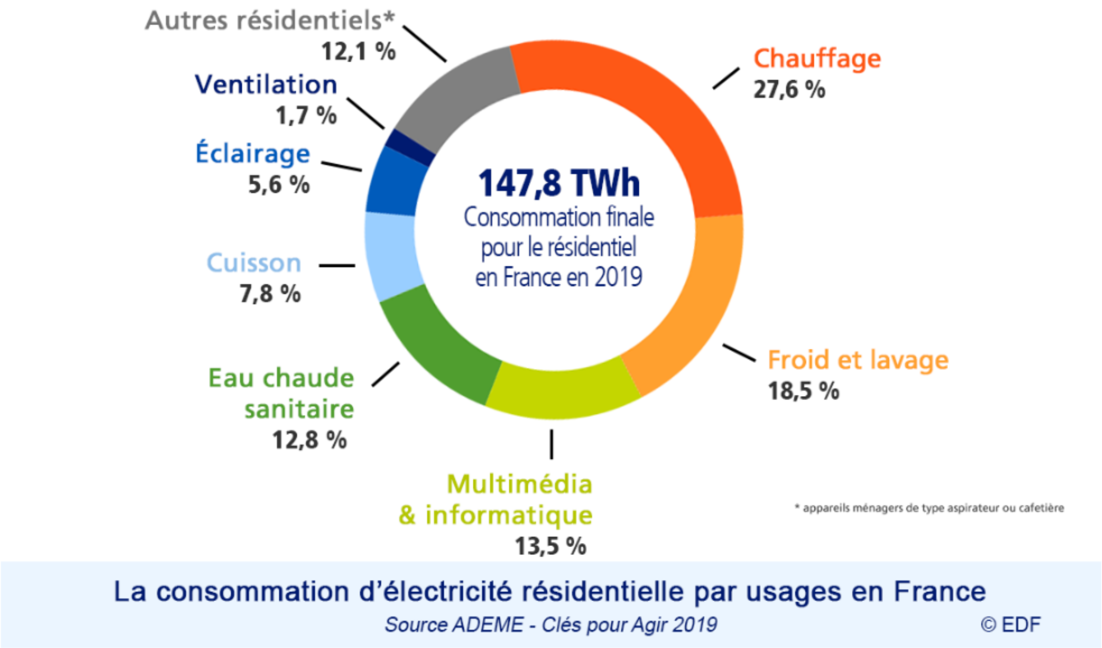


Figure 5.2: Repartition of the use of electricity at home in France in 2019.

With such a repartition, it can be easy to evaluate how the lockdown impacted the residential consumption. Indeed, since the lockdown happened during spring, the heating consumption should a priori not have been much impacted by a lockdown, which would not have been the case for a lockdown during winter. For other parts of consumption, one could juste take the usual consumption and multiply it to obtain the consumption when people are at home all the day. For example, the cooking consumption can be increased by taking into account the fact that one meal is added each day of lockdown[31]. Such a reflexion can be extended to each part of the residential consumption : for the multimedia and informatic part, the consumption could be increased by estimating the mean time that people usually spend on it and then estimating this time during lockdown. Obviously, all the ideas developed in this paragraph apply only to people who were concerned by the lockdown, not to people that had to go to work because their job was considered ”essential”.

Now that global ideas to evaluate the modifications of the consumption in the different sectors have been presented, the question of the duration of these modifications can be studied. Indeed, as shown in section 4.3.3, the drop during the year 2020 lasted longer than just the end of the lockdown. This can be explained by the measures taken by government and industries or companies : even if it was no longer required, many companies or industries kept applying working from home so the modifications explained here above for the primary and secondary sectors also kept applying. Another explanation of the recovering of a ”normal” consumption in September can be the complete reopening of schools. These are first ideas of explanation, to have a more precise and rigorous explanation, a specific research on the measures that were still applied or on the percentage of companies that continued working from home is needed.

All these ideas are just qualitative ideas because it would need a more rigorous approach to have precise figures of the modifications between normal period and lockdown period. The aim of this thesis was to have a global forecasting model able to include a shock like

---

[31] Usually people do not eat at home at noon but rather eat at work or school for example.

the one due to the COVID-19 pandemic but another work could focus on the precise evaluation of the modifications of electricity consumption due to the measures. Then, with precise modifications to include in the input signal of the model presented in this thesis, the model would be accurate and rigorous.

These ideas could also be used in another work on another type of shock (like the current war in Ukraine for example) : once the general model is validated, one should study how the event impacts the different sectors in the field of the study. To have an example that would impact the electricity production, since electricity is the focus of this thesis, one could imagine that the nuclear power plant with the highest production capacity of a country would be deactivated (due to a political decision or an accident for example) : this deactivation could result in a modification[32] of the production in all the other nuclear power plants that were working lower than their capacity, so an analysis of their capacity and recent production could lead to a clear determination of the modifications due to this hypothetical event.

Now that the question of the determination of the modifications due to the pandemic has ben qualitatively answered, the aim of this thesis begins to be fullfilled and it is time to evaluate the limitations and possible improvements that can be brought to this work. This is the aim of the section 6.

---

[32] In reality, one could expect to have a modification in other types of production than nuclear but this is just an example.

6 Limitations and Possible Improvements

The aim of this section is to present the limitations of the work done in this thesis and to give some possible improvements that could be done in another work.

The main limitation of this thesis is the analysis done in section 5 that is only qualitative and not quantitative. A precise analysis based on the ideas that were given in this section could really improve the rigor of this work. This has not been done here mainly because the required data to make such an analysis were out of reach but it can be the subject of another work.

In terms of accuracy of prediction, a first possible improvement would be to take an average temperature instead of the temperature of a single place, even if this place is placed at the center of the country as in this work. This is especially important in very large country as the USA or China for example, but it is not mandatory in France which can be considered as a not so large country compared to the ones cited just above. Indeed, an average temperature in France during the year 2020 has been computed, using the temperatures of 8 towns all along the country[33] and the figure 6.1 shows the comparison with the temperature of Clermond-Ferrand (since it is the one that has been used in this thesis). This figure shows that there is a difference in terms of values but the global trend is the same so if the difference is not too high it is not a reason of unaccuracy of the model. The mean difference all over 2020 is 1.62°C, which is not too high so the choice made in this work works well, but it should be kept in mind that taking an average value is the right way for a similar work in a large area.

---

[33] Bordeaux, Clermont-Ferrand, Grenoble, Lille, Nantes, Paris, Strasbourg and Toulouse.

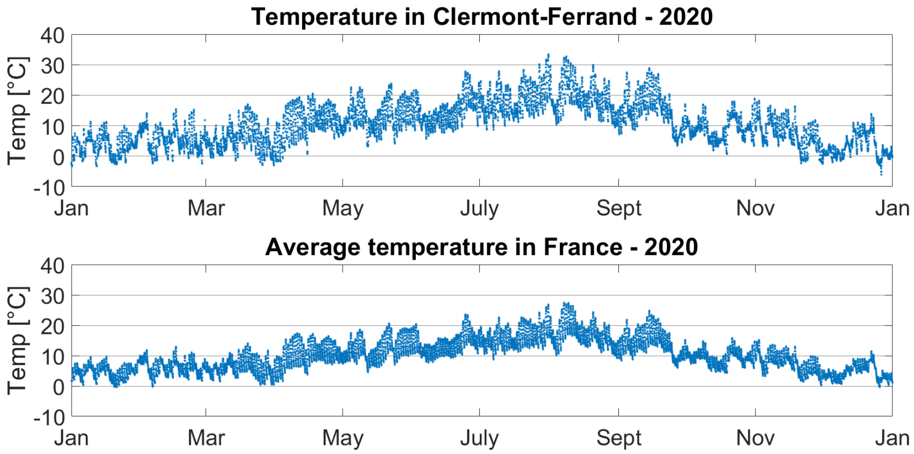


Figure 6.1: Temperature in Clermont-Ferrand and average temperature in France in 2020. The data are taken from [14] and the average has been computed using the data of 8 towns all along France.

Other possibilities of improvement are done in [15] and [32], two master thesis that use the general model presented in section 3 and that bring modifications to improve its accuracy. For example, [15] uses the model inversed, i.e. it uses the load to predict the temperature and the author proposes to use not only the current value of the electricity load but also some of the recent values. On the other hand, the author of [32] proposes for example not to use the hourly temperature but rather an average daily temperature for each hour of the day and to add only the relevant temperatures of the day : the three temperatures that lead to the highest load and the three temperatures that lead to the lowest load, all the other temperatures of the day being forgotten. These ideas are just examples taken from the two thesis and they lead to good accuracy, so they could also be used with the model presentedin this thesis to improve its accuracy.

Finally, the major possible improvement would probably to make the same model as in this thesis but to forecast a larger area, for example the whole Europe instead of just France, because this model could also apply for all the other countries in Europe.

7 Conclusion

The main objective of this thesis was to develop a forecasting model able to include a shock like the one due to the COVID-19 pandemic. This model had to be developed in order to make some analysis on the production of electricity in France in 2020.

For this, a first general forecasting model without taking a shock into account was presented in section 3. This model, based on the temperature and on calendar variables, has been proven in literature and in this thesis to have very accurate results (less than 4% of error) for a prediction without a shock. This model was modified in section 4 by using it as an input signal in an autoregressive model, autoregressive model which has also been proven in literature to have accurate results. With this autoregressive model, the shock due to the pandemic could be modeled with an error of about 6.4%, so the objective of developing an accurate enough model is fullfilled. With this model, it could also be seen that the lockdown imposed by the government happened at a good period of the year in terms of electricity production. Indeed, taking this model and placing the lockdown at different moments showed that it would have resulted in an even more important drop that the one that was effectively observed in 2020.

Although the developed model is validated and allows to make some analysis, section 5 discusses the rigor of the developing method and presents several ideas in order to improve this rigor. All the ideas presented in this section could be the subject of another work that would improve this one.

After the presentation of the limitations of this thesis and the possible improvements made in section 6, a conclusion of this thesis can be to invite a potential reader to improve the work done by using one of the ideas of improvement of section 6 or to develop those presented in section 5 in order to make a complete analysis.

References

[1] URL: `https://en.wikipedia.org/wiki/Energy_forecasting`.

[2] Rafał Weron. "Electricity price forecasting: A review of the state-of-the-art with a look into the future". In: *International Journal of Forecasting* 30 (2014), pp. 1030–1081.

[3] URL: `https://www.neuraldesigner.com/blog/electricity_demand_forecasting`.

[4] Peng Jiang, Yee Van Fan, and Jiři Jaromír Klemeš. "Impacts of COVID-19 on energy demand and consumption: Challenges, lessons and emerging opportunities". In: *Applied Energy* 285 (2021).

[5] Kevin M. Camp et al. *From the barrel to the pump: the impact of the COVID-19 pandemic on prices for petroleum products.* Tech. rep. U.S. Bureau of Labor Statistics, Oct. 2020.

[6] URL: `https://transparency.entsoe.eu/`.

[7] Tao Hong. "Short Term Electric Load Forecasting". PhD thesis. University of North Carolina at Charlotte, 2010.

[8] URL: `https://en.wikipedia.org/wiki/Global_Energy_Forecasting_Competition`.

[9] URL: `https://www.kaggle.com/competitions/global-energy-forecasting-competition-2012-load-forecasting/overview/description`.

[10] Tao Hong et al. "Probabilistic energy forecasting : Global Energy Forecasting Competition 2014 and beyond". In: *International Journal of Forecasting* 32 (2016), pp. 896–913.

[11] URL: `https://otexts.com/fpp2/useful-predictors.html`.

[12] URL: https://www.analyticsvidhya.com/blog/2015/11/easy-methods-deal-categorical-variables-predictive-modeling/.

[13] URL: https://bookdown.org/egarpor/PM-UC3M/lm-ii-qualpred.html.

[14] URL: https://re.jrc.ec.europa.eu/pvg_tools/en/.

[15] Masoud Sobhani. "A load-based temperature prediction model for anomaly detection". MA thesis. University of North Carolina at Charlotte, 2017.

[16] URL: https://en.wikipedia.org/wiki/Cross-validation_(statistics).

[17] URL: https://www.investopedia.com/terms/e/economic-shock.asp#:~:text=Economic%5C%20shocks%5C%20are%5C%20random%5C%2C%5C%20unpredictable,influence%5C%20either%5C%20supply%5C%20or%5C%20demand..

[18] Aviad Navon et al. "Effects of the COVID-19 Pandemic on Energy Systems and Electric Power Grids—A Review of the Challenges Ahead". In: *Energies* (2021).

[19] Hassan Qudrat-Ullah. "Understanding the dynamics of electricity generation capacity in Canada : A system dynamic approach". In: *Energy* 59 (2013), pp. 285–294.

[20] Mohamd Laimon et al. "Energy Sector Development: System Dynamics Analysis". In: *Applied Sciences* 10.1 (2020).

[21] URL: https://www.researchgate.net/profile/Nidal-Kamel-2.

[22] Nidal S. Kamel and Zuhairi Baharudin. "Autoregressive method in short term load forecast". In: *2008 IEEE 2nd International Power and Energy Conference* (2008), pp. 1603–1608.

[23] Nidal S. Kamel and Zuhairi Baharudin. "One Week Ahead Short Term Load Forecasting". In: *Inproceedings ofthe 7th IASTED International Conference on Power and Energy Systems* (2007), pp. 91–96.

[24] Nidal S. Kamel and Zuhairi Baharudin. "Short term load forecast using Burg autoregressive technique". In: *International Conference on Intelligent and Advanced Systems 2007 (ICIAS 2007)* (2007), pp. 912–916.

[25] URL: `https://en.wikipedia.org/wiki/State-space_representation#:~:text=In%5C%20control%5C%20engineering%5C%2C%5C%20a%5C%20state,differential%5C%20equations%5C%20or%5C%20difference%5C%20equations.`.

[26] M. Gevers and L. Vandendorpe. *INMA 1731 : Processus stochastiques, estimation et prédiction*. Tech. rep. 2011.

[27] Peter J. Brockwell and Richard A. Davis. *Introduction to Time-Series and Forecasting*. Springer, 2002.

[28] URL: `https://fr.wikipedia.org/wiki/Chronologie_de_la_pand%5C%C3%5C%A9mie_de_Covid-19_en_France`.

[29] URL: `https://en.wikipedia.org/wiki/Economic_sector`.

[30] URL: `https://en.wikipedia.org/wiki/Tertiary_sector_of_the_economy`.

[31] URL: `https://www.edf.fr/groupe-edf/espaces-dedies/l-energie-de-a-a-z/tout-sur-l-energie/l-electricite-au-quotidien/la-consommation-d-electricite-en-chiffres`.

[32] Shreyashi Shukla. "Daily Load Forecasting with Hourly Temperatures". MA thesis. University of North Carolina at Charlotte, 2018.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
École polytechnique de Louvain
Rue Archimède, 1 bte L6.11.01, 1348 Louvain-la-Neuve, Belgique | **www.uclouvain.be/epl**