Project Data Mining M2: Wind Turbines in France: A Retrospective Analysis

Anna Blanpied, Maria Bracque Vendrell, Matthieu Donati Arranz

Semester 1 - 2023

Contents

1 Introduction

The use of wind energy can be seen as far back as 3000 years. Before the Industrial Revolution, wind energy was primarily harnessed through the use of windmills which directly converted wind into mechanical rotational energy. With the Industrial Revolution came the widespread use of coal and other fossil fuels that largely replaced wind energy as fossil fuels provided a more constant, reliable, and greater source of energy.

Despite the greatness of fossil fuels, they have a large negative environmental impact, particularly through atmospheric carbon emissions which lead to global warming. Therefore, over the past couple of decades, following the Kyoto Protocol (Japan, 1997), governments and social movements have been trying to reverse the trend by intensifying the use of renewable energies.

With the creation of generators, wind energy has emerged as one of the most promising technologies to address global challenges related to sustainable energy supply, greenhouse gas emissions reduction, and the transition to renewable energy sources. However, wind turbines are a very controversial topic due to their visual and sound impact.

The objective of this project is to give an overview of the history of wind turbines in France. To perform a data analysis on their performance between 2013 and 2022. Finally, to create a model to evaluate the power produced per wind farm throughout the years by using a wind turbine power calculation equation and the wind power of the weather stations across France.

To do so, we will first provide a historical overview of wind turbines in France. Then, we will delve into an analysis of both power capacity and generation from these turbines. We will finally focus on the modeling phase, where we estimate daily power generation throughout 2021. To accomplish this, we initiated a wind model to predict wind speeds at each turbine park. We applied these findings, associated with the literature, to estimate power generation. The results we derived indicate that while optimistic, our model's fluctuations closely align with real-world values.

2 Overview of Wind Turbines in France

2.1 Historical Overview of Wind Turbines in France

The history of wind energy in France started with the introduction of the Darrieus turbine. The Darrieus turbine is a vertical-axis wind turbine, conceived by French inventor George Darrieus in 1931. Since then, a total of 20 916 MW (as date of December 31, 2022) of wind power capacity have been installed in France.

By the end of 2015, the total onshore installed capacity of 10,358 MW consisted of 5,956 turbines. According to figures released by the French government, there were nearly 2,000 wind farm installations across France by the end of 2019. While France has been a relatively late developer in wind power compared to other European countries, it has set the target of more than doubling onshore wind power capacity from 2015 levels by 2023. In February 2022, President Emmanuel Macron declared France's commitment to constructing 50 offshore wind farms, aiming for a cumulative capacity of at least 40 GW by 2050.

Additionally, the existing framework supporting renewable energy sources in France is based on Law No. 2015-992 dated 17 August 2015, titled "On Energy Transition for Green Growth". This legislation sets ambitious national energy goals. The PPE (Pluriannual Energy Program), updated by Decree No. 2020-456 on 21 April 2020, serves as a strategic roadmap for the French Government's energy initiatives over the next ten years. The primary objective is to guide France towards becoming a carbon-neutral nation by 2050. In 2023, according to the Ministère de la Transition Ecologique, renewable energies account for 20.7% of gross final energy consumption, and among renewable energies, wind power accounts for 11%.

2.2 Analysis of Wind Turbines in France

2.2.1 Wind Turbine Capacity

Data

Our data set contains all the terrestrial wind turbines in France, active or ordered. This data set is publicly available on a government website, Géorisques.gouv. The data set contains 11759 observations. We only kept the active windmills, which represent 7292 observations. There are in total 24 variables:

Variable	Description
id_aerogenerateur	Unique identifier for the wind turbine
id_parc	Unique identifier for the wind turbine park
code_insee	Code for the commune
nom_commune	Name of the commune
code_dept	Code for the department
code_reg	Code for the region
puissance	Power of the wind turbine
hauteur_totale	Total height of the wind turbine
hauteur_mat_nacelle	Height from base to nacelle
diametre_rotor	Rotor diameter
cote_ngf	Altitude above sea level
periode_allumage_lib	Lighting period (free text)
periode_allumage_desc	Description of the lighting period
type_feu_lib	Type of lighting (free text)
type_feu_desc	Description of the lighting type
date_mise_en_service	Date of commissioning
constructeur	Manufacturer of the wind turbine
reference_modele	Model reference
x_aerogenerateur	X-coordinate of the wind turbine
y_aerogenerateur	Y-coordinate of the wind turbine
epsg	EPSG code
libelle	Label
date_maj	Last update date
nom_eolienne	Name of the wind turbine

Table 1: Variables of the dataset on terrestrial wind turbines in France

We created four additional variables: *année* (the year the windmill was built), *latitude*, and *longitude* (the GPS coordinates), and *Région* (the region where the windmill is located).

Some observations required rectification as they had a power of 80MW, which seemed unusually high. In such cases, we conducted cross-referencing with other observations to obtain the correct values. To achieve this, we compared these suspicious windmills to others of the same model. If the model had only one power value in the data set, we assigned this value to the windmill. If there were multiple values, we first checked if there was another windmill of the same model from the same park. In such cases, we assigned the power value of the identified windmill to our suspicious windmill. If no such windmill was found, we examined windmills with similar characteristics (such as height, rotor diameter, and nacelle height) of the same model.

A few interesting graphs

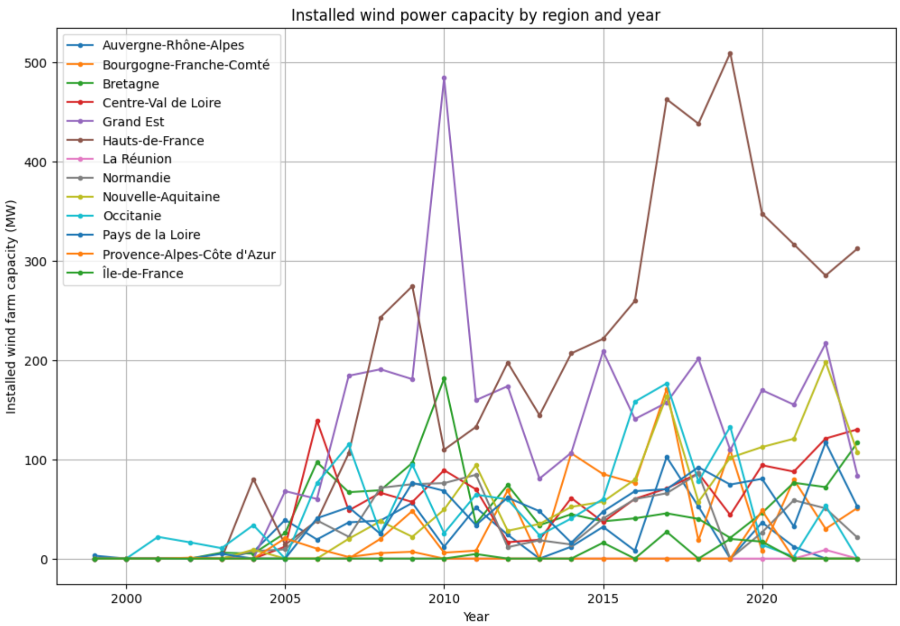


Figure 1: Installed wind power capacity by region and year

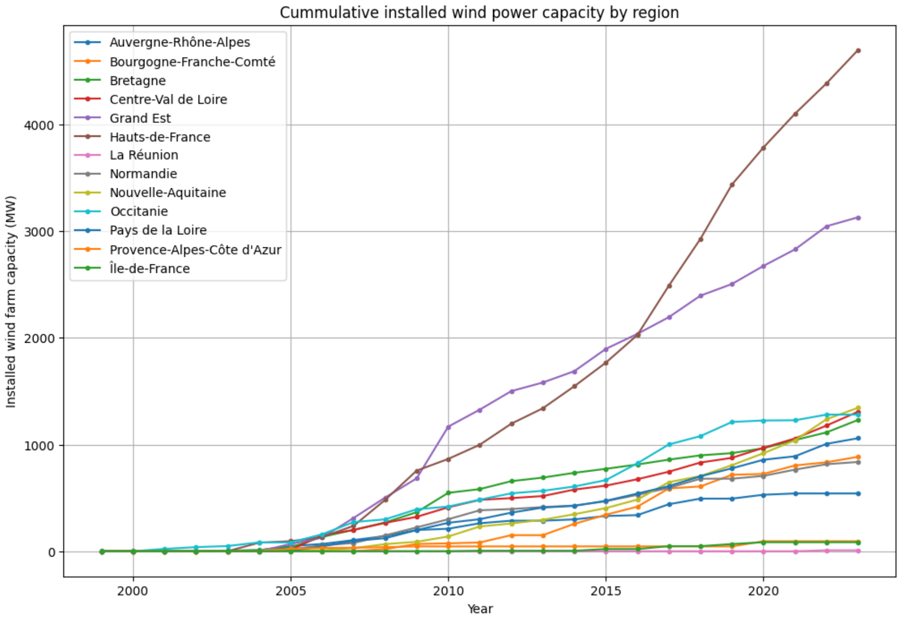


Figure 2: Cumulative installed wind power capacity by region

We can see that most of the regions have linear trends apart from "Haut-de-France" which has more of an exponential trend.

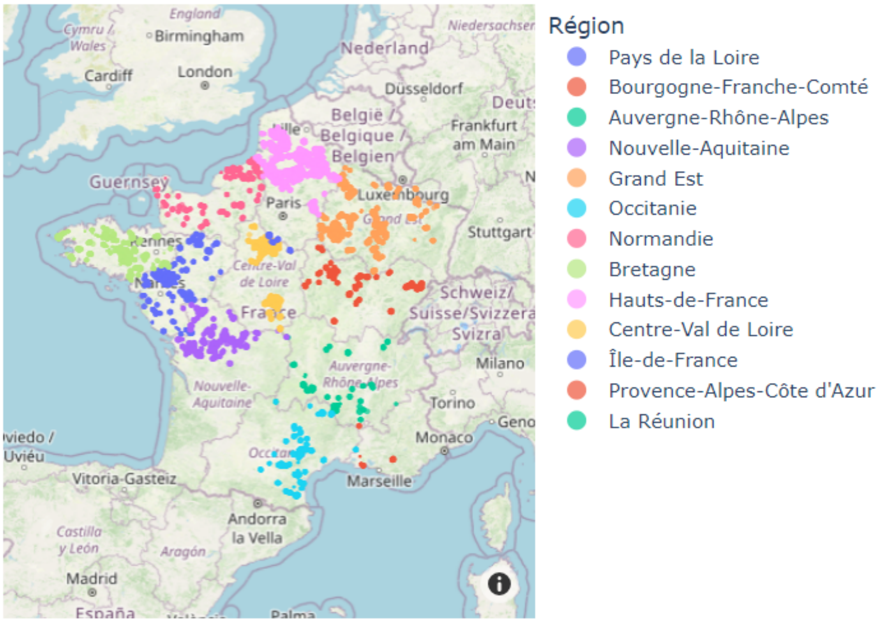


Figure 3: Map of the different wind farms in France

2.2.2 Wind Power Generation

Data

We use a data set from ODRÉ (OpenData Réseaux-Energies) for this section. This data set contains regional data from January 2013 from the éCO2mix application. éCO2mix is an easy-to-use tool created by RTE to help consumers better understand and consume electricity. It provides all electricity consumption and production indicators in real-time, 24 hours a day, at a national and regional level. Among other electricity production methods, it contains the level of energy production by wind turbines every 30 minutes by region.

Data cleaning

First, we deleted the columns that don't relate to wind turbines. After, we converted the 'Date' variable to a DateTime variable, and created three new columns with the month, the year, and the month and year that will be helpful in the future analysis. Finally, we deleted the 108 observations with missing values since we have 1 980 288 observations. The final variables of this data set are thus:

Variable	Description
Code INSEE région	Code for the commune
Région	Name of the region
Date	Energy production date
Heure	Energy production hour
Date - Heure	Energy production date and hour
Eolien (MW)	Energy production
Year	Energy production year
Month	Energy production month
month_year	Energy production year and month

Table 2: Variables of the data set on wind energy production in France

Data Analysis

First, let us have a look at the energy production of the different wind parks per region.

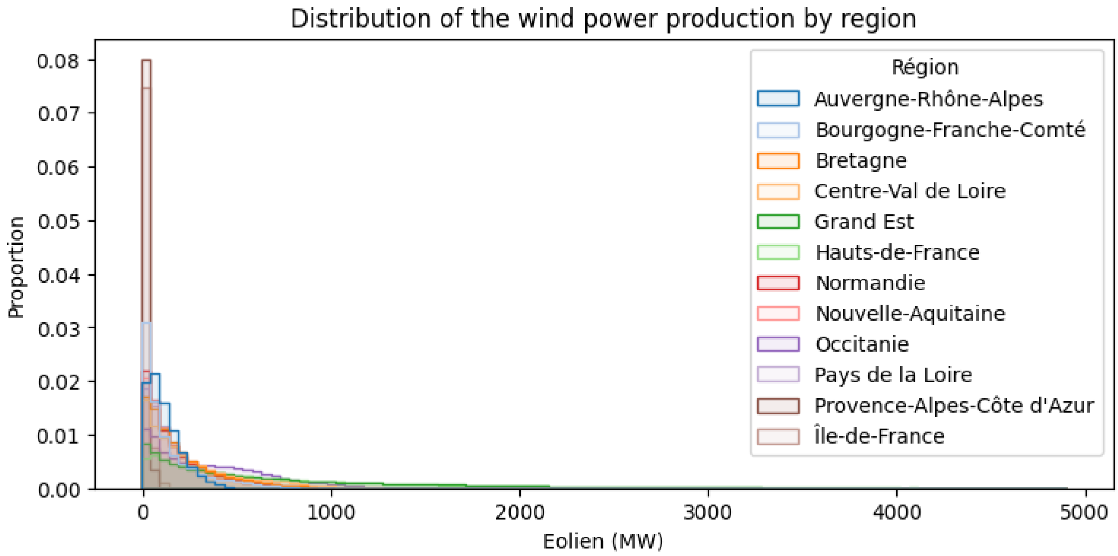


Figure 4: Distribution of the wind power production by region

We notice that overall, wind power generation follows a power law in each region. This means that the majority of observations are low production values, while high production values are rare. This could be due to various factors, such as the variability of weather conditions (e.g. wind speed and direction) that affect wind power generation. In addition, wind power generation capacity may vary according to the technology and infrastructure available in each region.

Then, we looked at the average production per day, month, and year for each region.

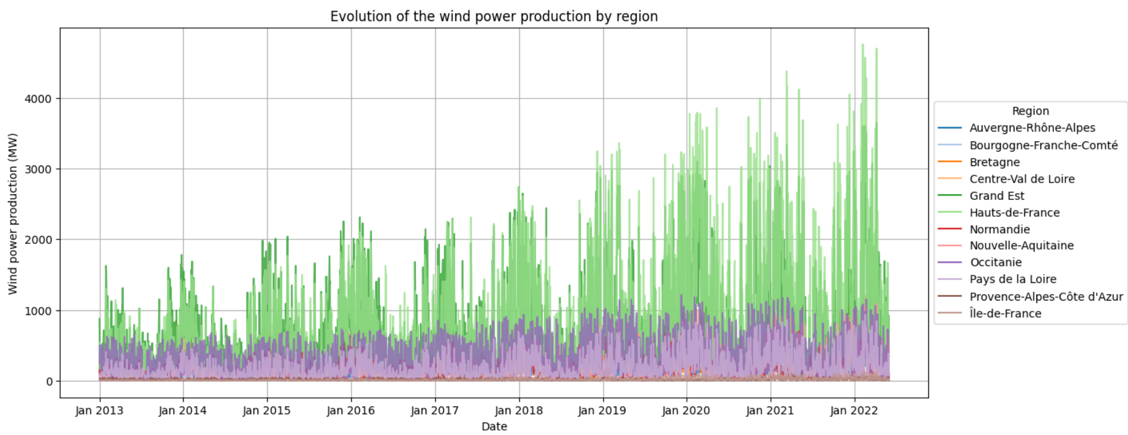


Figure 5: Evolution of the wind power production by region

We notice that it is a little hard to see all the regions on the same graph. However, we still notice that there is a kind of cyclicity that depends on the months of the year. This can also be seen in the following graph when we do not distinguish between regions.

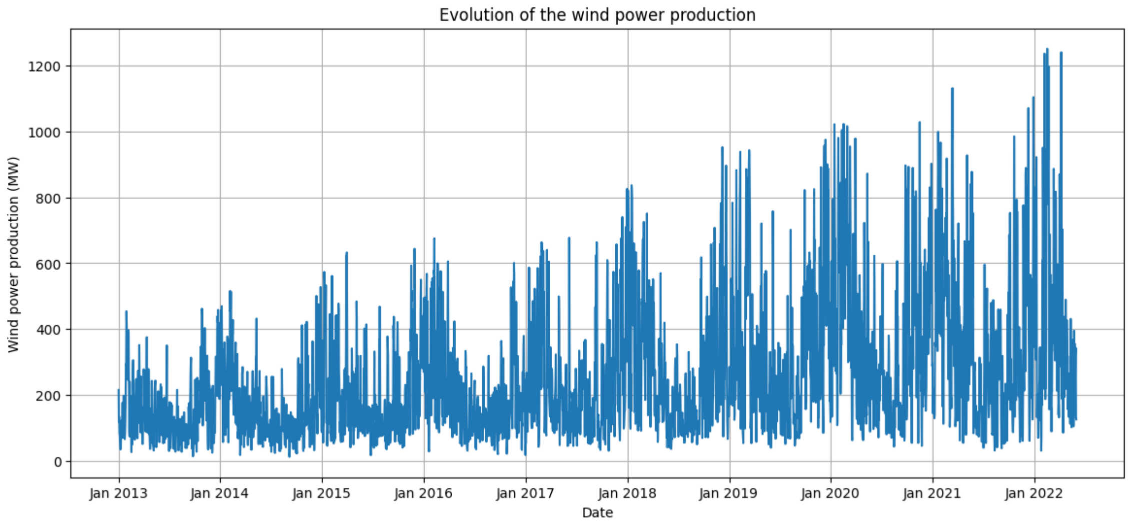


Figure 6: Evolution of the wind power production

To better visualize the cyclicity, and to take into account the difference in capacity between wind farms and years in each region, we're going to normalize the data. To do this, we need the capacity of wind turbines per region and year, we used the dataset from the previous section.

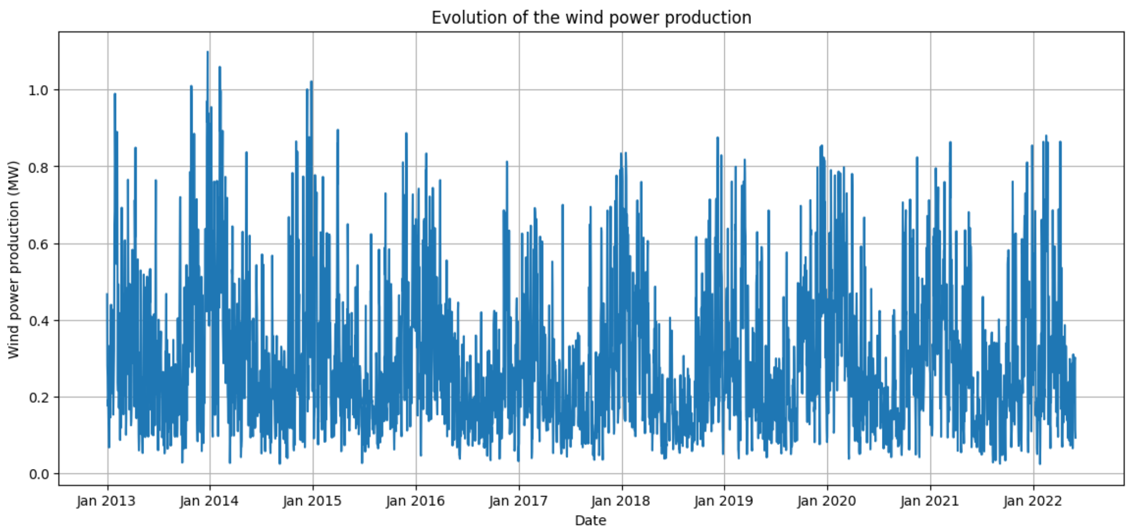


Figure 7: Evolution of the wind power production

We see that there is a cyclicity that depends on the months of the year. The production is higher during winter and lower during summer.
Thus, for the modelization part, we are going to focus on predicting the production per wind farm for only one year.

3 Modelisation

The goal of this part is to predict the electricity generated by wind turbines (individually or by park). As there was no available dataset containing this specific information, we decided to generate it ourselves. Our approach involved predicting the wind speed at each wind turbine's location at the desired moment for analysis. By accurately estimating this wind speed, we could then calculate the power output of the turbines using literature, allowing us to determine the anticipated electricity production.

3.1 The Wind

Wind plays a crucial role in assessing the performance of wind turbines. It is one of the main factors in calculating the power generated by these turbines, thereby providing insight into the anticipated electricity output. Our primary goal was to get the wind speed at each wind turbine park location. Given the absence of available data specific to these locations, we generated this information by leveraging wind speed data collected from nearby weather stations.

3.2 Data

The data we used for the wind comes from international surface observation reports (SYNOP) transmitted through the World Meteorological Organization's global telecommunication system (SMT). These observations encompass various atmospheric conditions measured, such as temperature, humidity, wind direction and speed, atmospheric pressure, precipitation levels, and observations like weather conditions, cloud descriptions, and visibility from the Earth's surface. Depending on the instruments and local specifications, additional parameters such as snow depth, ground conditions, and more might also be available.

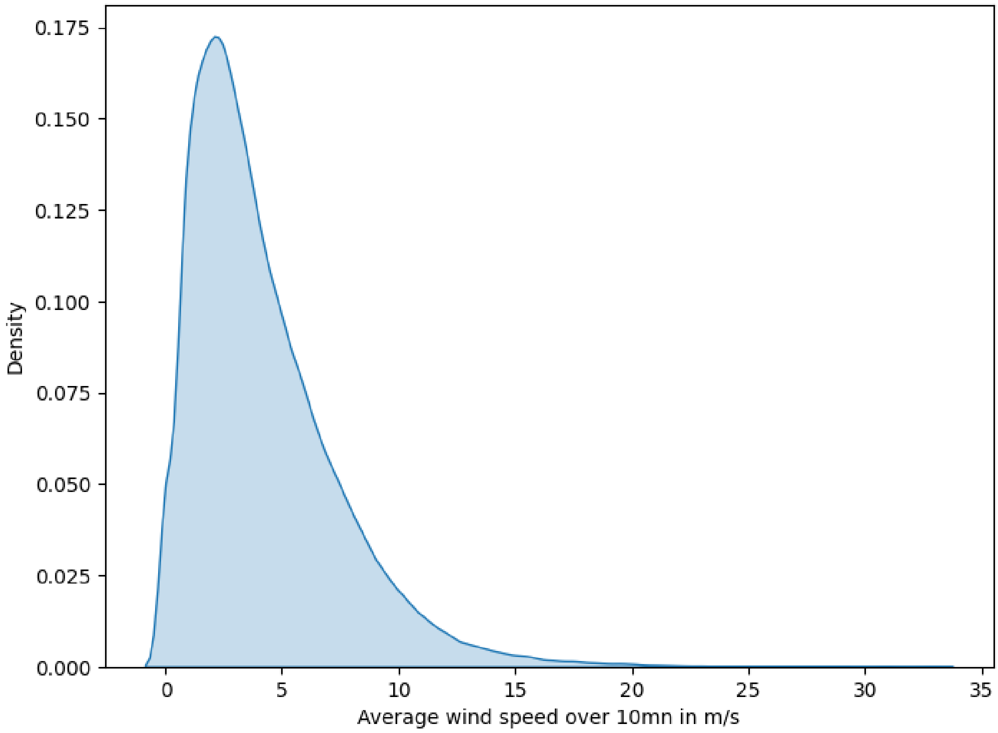
We extracted a dataset from SYNOP containing the data for the year 2021. Appendix 1 includes a comprehensive list of all variables within this dataset. For our study, we selected the following variables from this dataset:

- ID OMM station
- **Location variables**: communes (name), communes (code), EPCI (name), EPCI (code), department (code), department (name), region (name), region (code)
- **Time variables**: Date
- **Geographical variables**: Latitude, Longitude, Altitude
- Vitesse du vent moyen 10 mn

Then, we conducted feature engineering to enhance our dataset's information. First, we decomposed the *Date* variable into four distinct time variables: *Year*, *Month*, *Day*, and *Hour*. This step allowed us to incorporate temporal aspects into our model effectively. Additionally, using the Haversine distance formula, which uses the longitude and latitude of each location, we computed the distances between all stations in kilometers, enriching our dataset with spatial relationships.

3.3 Modelisation

The target variable under analysis is the average wind speed over 10 minutes (*Vitesse du vent moyen 10 mn*), measured in meters per second (m/s). This variable represents the average wind speed observed over a 10-minute interval and exhibits the following density distribution:



To predict this variable, our initial approach involved employing a LightGBM model. To establish a benchmark, we started by executing the model without any specific tuning or parameter adjustments. This preliminary run provided us with the following set of results:

Metric	**Value**
MAE of the model	1.649
RMSE of the model	2.291
R^2 of the model	49.79%

Table 3: Model Evaluation Metrics

Given the initial results were not satisfying, we opted to conduct a more comprehensive optimization approach. To enhance the model's performance, we executed a grid search, systematically exploring various parameter combinations within the

LightGBM framework. This yielded a set of optimized parameters that maximized the efficiency of our model. The resulting best parameters obtained from this grid search are as follows:

- learning_rate = 0.1
- max_depth = 15
- n_estimators = 300
- num_leaves = 2187

and the model's metrics are the following :

Metric	**Value**
MAE of the model	1.191
RMSE of the model	1.732
R^2 of the model	71.32%

Table 4: Model Evaluation Metrics

Given the improved performance of this refined model, we chose to use it for predicting the wind speeds across the wind turbine parks.

3.4 Application

Now that we have produced a model, we will predict the wind for the year 2021 at the different wind turbines. With this prediction, we are going to be able to compute the theoretical capacity in kilowatts of each wind turbine in 2021 with the following formula:

$$P = 1/2 * \rho * S * V$$

where P is the capacity in watts, ρ is the air density which is approximately 1.293 kg/m3 at 0 °C and 1.204 kg/m3 at 20 °C, S is the area of a circle with a radius equal to the length of a blade, and V the wind speed in meters per second.[1] We obtain the final graph:

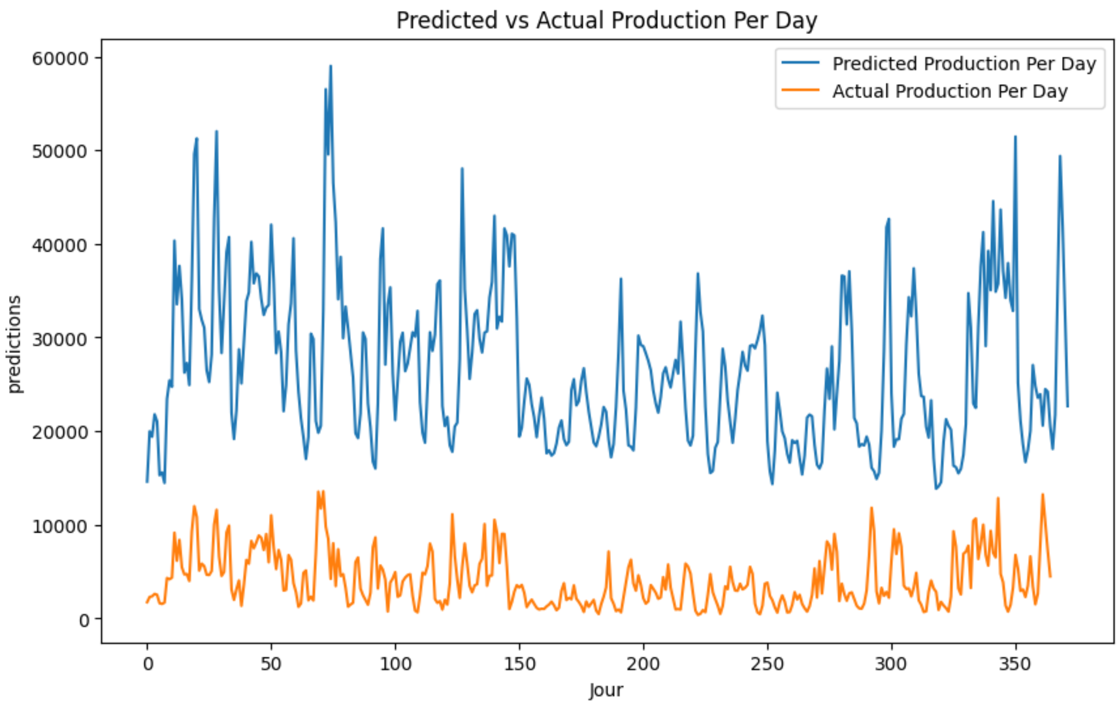


Figure 8: Predicted vs Actual Production Per Day

We can see that the predictions are way too optimistic. However, we can also see that the fluctuations are very similar. Thus we can conclude that even though our prediction model did not manage to predict the real production, we managed to capture pretty correctly the seasonal fluctuations.

1. source: https://plum.fr/blog/energie-ecologie/rendement-eolienne/

4 Conclusion

Overall, in our analysis, we saw the evolution at regional and national levels of wind turbine capacity.

During the analysis, we noticed two different trends. First, regarding wind power generation, we observed fluctuations, emphasizing the seasonal and cyclical nature of wind energy. Indeed, we noticed that more energy is produced in winter. Second, there is an overall increasing trend over the years, mirroring the capacity analysis where we saw that more and more wind turbines have been installed over the years.

Then, in the prediction part of our report, we tried to predict wind speeds using a LightGBM model and weather station data. This allowed us to estimate the electricity production of wind turbines by applying a theoretical model to these wind speed predictions. While our model shows promise in capturing seasonal patterns, predictions overestimate production. There is room for improvement in accurately predicting actual production.

This analysis provides a good foundation and starting point for future studies, offering opportunities to refine the prediction model and stay up to date on what is happening with wind energy. Looking ahead, doing more research on where to place the wind farms could help make wind energy work better. This is important for France's goal of achieving carbon neutrality by 2050 and sustainable energy practices.

Bibliography

"Chiffres clés des énergies renouvelables - Édition 2023." 2023. Ministère de la Transition écologique. Accessed December 21, 2023. https://www.statistiques.developpement-durable.gouv.fr/edition-numerique/chiffres-cles-energies-renouvelables-2023/livre?page=4.

Gipe, Paul, and Erik Möllerström. 2022. "An overview of the history of wind turbine development: Part I—The early wind turbines until the 1960s." *Wind Engineering* 46 (6): 1973–2004. https://doi.org/10.1177/0309524X221117825. https://doi.org/10.1177/0309524X221117825.

Government, French. 2015. "Law No. 2015-992 of 17 August 2015." Accessed December 21, 2023. https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000031044385.

———. 2019. "French wind energy statistics." Accessed December 21, 2023. https://www.statistiques.developpement-durable.gouv.fr/energie-climat/r/energies-renouvelables.html?tx_ttnews%5Btt_news%5D=41270&cHash=c9d0b3fd55a36c0a8787acbdc2272a3a.

———. 2020. "Decree No. 2020-456 of 21 April 2020." Accessed December 21, 2023. https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000041788175.

———. 2023. "French wind energy report." Accessed December 21, 2023. https://www.gouvernement.fr/en/french-wind-energy-report.

l'Eolien, Le Journal de. n.d. "Chronologie de l'éolien en France." (accessed: 13.11.2023). https://www.journal-eolien.org/tout-sur-l-eolien/chronologie-de-l-eolien-en-france/.

Wikipedia. 2021a. "Darrieus wind turbine." Accessed December 21, 2023. https://en.wikipedia.org/wiki/Darrieus_wind_turbine.

Wikipedia. 2021b. "Wind power in France." Accessed December 21, 2023. https://en.wikipedia.org/wiki/Wind_power_in_France.

5 Appendix

Appendix 1: Variables of the SYNOP dataset

Variable	Description
ID OMM station	Identification code for the station
Date	Date of observation
Pression au niveau mer	Sea level pressure
Variation de pression en 3 heures	Pressure change in the last 3 hours
Type de tendance barométrique	Type of barometric tendency
Direction du vent moyen 10 mn	Average wind direction over 10 minutes
Vitesse du vent moyen 10 mn	Average wind speed over 10 minutes
Température	Temperature
Point de rosée	Dew point
Humidité	Humidity
Visibilité horizontale	Horizontal visibility
Temps présent	Present weather
Temps passé 1	Past weather 1
Temps passé 2	Past weather 2
Nebulosité totale	Total cloud cover
Nébulosité des nuages de l'étage inférieur	Cloud cover of low-level clouds
Hauteur de la base des nuages de l'étage inférieur	Height of base of low-level clouds
Type des nuages de l'étage inférieur	Type of low-level clouds
Type des nuages de l'étage moyen	Type of mid-level clouds
Type des nuages de l'étage supérieur	Type of high-level clouds
Pression station	Station pressure
Niveau barométrique	Barometric altitude
Géopotentiel	Geopotential
Variation de pression en 24 heures	Pressure change in the last 24 hours
Température minimale sur 12 heures	Minimum temperature over 12 hours
Température minimale sur 24 heures	Minimum temperature over 24 hours
Température maximale sur 12 heures	Maximum temperature over 12 hours
Température maximale sur 24 heures	Maximum temperature over 24 hours
Température minimale du sol sur 12 heures	Minimum ground temperature over 12 hours
Méthode de mesure Température du thermomètre mouillé	Method of measuring wet-bulb temperature
Température du thermomètre mouillé	Wet-bulb temperature
Rafale sur les 10 dernières minutes	Wind gusts in the last 10 minutes
Rafales sur une période	Wind gusts over a period
Période de mesure de la rafale	Period of wind gust measurement
État du sol	Ground state
Hauteur totale de la couche de neige, glace, autre au sol	Total height of snow, ice, or other on the ground
Hauteur de la neige fraîche	Height of fresh snow
Période de mesure de la neige fraîche	Period of fresh snow measurement
Précipitations dans la dernière heure	Precipitation in the last hour
Précipitations dans les 3 dernières heures	Precipitation in the last 3 hours
Précipitations dans les 6 dernières heures	Precipitation in the last 6 hours
Précipitations dans les 12 dernières heures	Precipitation in the last 12 hours
Précipitations dans les 24 dernières heures	Precipitation in the last 24 hours
Phénomène spécial 1	Special phenomenon 1
Phénomène spécial 2	Special phenomenon 2
Phénomène spécial 3	Special phenomenon 3
Phénomène spécial 4	Special phenomenon 4
Nébulosité couche nuageuse 1	Cloud cover layer 1
Type nuage 1	Cloud type 1
Hauteur de base 1	Base height 1
Nébulosité couche nuageuse 2	Cloud cover layer 2
Type nuage 2	Cloud type 2
Hauteur de base 2	Base height 2

Variable	**Description**
Nébulosité couche nuageuse 3	Cloud cover layer 3
Type nuage 3	Cloud type 3
Hauteur de base 3	Base height 3
Nébulosité couche nuageuse 4	Cloud cover layer 4
Type nuage 4	Cloud type 4
Hauteur de base 4	Base height 4
Coordonnées	Coordinates
Nom	Name
Type de tendance barométrique.1	Type of barometric tendency 1
Temps passé 1.1	Past weather 1.1
Temps présent.1	Present weather 1
Température (°C)	Temperature (°C)
Température minimale sur 12 heures (°C)	Minimum temperature over 12 hours (°C)
Température minimale sur 24 heures (°C)	Minimum temperature over 24 hours (°C)
Température maximale sur 12 heures (°C)	Maximum temperature over 12 hours (°C)
Température maximale sur 24 heures (°C)	Maximum temperature over 24 hours (°C)
Température minimale du sol sur 12 heures (en °C)	Minimum ground temperature over 12 hours (in °C)
Latitude	Latitude
Longitude	Longitude
Altitude	Altitude
communes (name)	Name of the locality
communes (code)	Code of the locality
EPCI (name)	Name of the Public Establishment of Inter-Municipal Cooperation
EPCI (code)	Code of the Public Establishment of Inter-Municipal Cooperation
department (name)	Name of the department
department (code)	Code of the department
region (name)	Name of the region
region (code)	Code of the region
$mois\_de\_l_{a}nnee$	Month of the year