# Climate-related renewable energy sources and carbon emissions: a machine learning-based investigation of electricity production in France

Mohammadehsan Eslahi[a,b,c,*]

[a]IESEG School of Management, UMR 9221 - LEM - Lille Economie Management, F-59000 Lille, France
[b]Univ. Lille, UMR 9221 - LEM - Lille Economie Management, F-59000 Lille, Franc
[c]CNRS, UMR 9221 - LEM - Lille Economie Management, F-59000 Lille, France

**Abstract**

This study characterizes the predictive relationship between electricity generation from climate-related renewable energy (CRE) sources (wind, solar photovoltaics, and small-scale run-of-river hydroelectricity) and daily $CO_2$ emissions in the French power sector over the 2013-2021 period. The results demonstrate that run-of-river hydroelectricity was the most important source among the three for predicting emissions, followed by wind energy. Empirical findings based on a counterfactual analysis also reveal that an increase in the share of electricity from CRE sources would have been associated with a statistically significant decrease in predicted emissions over the study period. Identification of the optimal mix of CRE sources for minimizing predicted emissions under four counterfactual scenarios of increased CRE production reaffirms the greater relative share of run-of-river hydroelectricity and wind energy within the mix. The findings provide fresh quantifiable evidence on the (relative) importance of CRE sources for carbon emissions reduction in the electricity sector in France.

*Keywords:* Climate-related Renewable Energy, Carbon Emissions, Electricity Production, Machine Learning

## 1. Introduction

Carbon dioxide ($CO_2$) emissions from energy use are one of the principal contributors to global climate change and represent almost 75% of all anthropogenic greenhouse gas emissions in the European Union (European Commission, 2021). Such emissions come under the influence of multiple climatic, social and economic factors, and are mainly rooted in the combustion of fossil fuels

*Corresponding author
*Email address:* m.eslahi@ieseg.fr (Mohammadehsan Eslahi)

for electricity production, transportation, industrial and agricultural purposes. Notwithstanding the fact that all sectors should be held accountable for reducing $CO_2$ emissions, the electricity sector is expected to play the lead role in the decarbonization of economy owing to its more pronounced ability to lower emissions in a cost and time-effective manner (Edenhofer, 2015; Rodrigues et al., 2020; Goh et al., 2018; Karmellos et al., 2016). Development of renewable energy share of electricity production, leveraging carbon capture and sequestration (CCS) technologies in power plants, and increasing nuclear energy supply are three alternative methods for mitigating $CO_2$ emissions in the electricity sector (Brouwer et al., 2016). Among these options, renewable energy sources have been argued to be the keystone of $CO_2$ mitigation (Rogelj et al., 2018), not only from an environmental point of view but also in the light of economic, social and political considerations (Waisman et al., 2019).

Being a pioneer in the battle against global warming, the European Union already has a significantly lower emissions intensity of electric power generation than other large economies such as the United States, Japan, China, India and Australia (International Energy Agency (IEA), 2020). The French electricity sector is comparatively even more decarbonized (Shirizadeh & Quirion, 2021), largely due to the considerable share of nuclear energy generation.[1] The 2020 report of the electricity transmission system operator of France[2] asserts that, in 2019, emissions from electricity production in the country reached approximately 21.16 million tonnes of $CO_2$ equivalent ($CO_2$-eq), accounting for 4.8% of total emissions (Réseau de Transport d'Électricité (RTE), 2020).

In alignment with France's carbon neutrality by 2050 objective set by the National Assembly in 2019 under the title "Ecological Emergency and Climate Crisis" (Ministère de la Transition écologique, 2019), $CO_2$ emissions from electricity production need to be further reduced either by retrofitting CCS to existing power plants, building new nuclear reactors or increasing the share of renewable electricity (Débat national sur la transition énergétique, 2013; Shirizadeh & Quirion, 2021). Two analyses conducted by the French Environment and Energy Management Agency (ADEME) have shown that the development of a new generation of nuclear energy would not be economically efficient for the French electricity system, and that in an ideal scenario, electricity generation from renewable sources would constitute the largest share–up to 95%–of electricity generation in France over the next few decades (see ADEME, 2015, 2018). It is in this regard that the nexus between $CO_2$ emissions in the electric power sector and various forms of renewable electricity generation needs careful analysis.

Figure 1 depicts the share of total electric energy produced by main fuel categories in France from 2013 to 2020. From this figure, two main observations are clear. First, while the share of the fossil fuel mix (i.e. fuel oil, coal and

[1] France has the world's largest share of electricity production by nuclear power, with about 70% of its electricity being generated from nuclear energy (World Nuclear Association, 2022).

[2] RTE–Réseau de Transport d'Électricité (`https://www.rte-france.com/`)

gas in aggregate) in total electricity production has only slightly decreased over this period (from 7.91% in 2013 to 7.51% in 2020), the shares of total electricity production from fuel oil and coal have decreased, respectively, from 0.69% and 3.61% in 2013 to 0.34% and 0.28% in 2020. This decrease has been accompanied by an increase in the share of total electricity production from gas (from 3.61% in 2013 to 6.89% in 2020), indicating a shift in the fossil fuel mix. Second, the share of the wind-solar-hydroelectric energy mix in total electricity production has increased from 17.44% in 2013 to 23.55% in 2020. Accompanied by a decrease in the share of total electricity produced from nuclear energy (from 73.34% in 2013 to 66.99% in 2020), this latter signifies a change in the composition of non-fossil fuel mix, i.e. a gradual transition from nuclear to wind-solar-hydro electricity.[3]

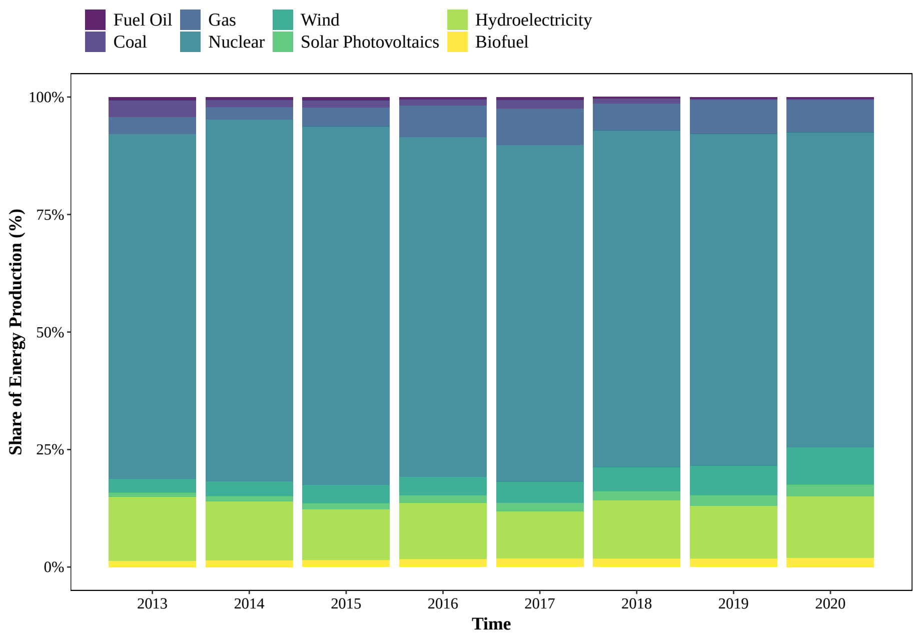


Figure 1: Share of total electric energy produced by main fuel categories in France from 2013 to 2020 (Data source: `https://opendata.reseaux-energies.fr/`).

The relationship between $CO_2$ emissions and renewable energy production and consumption has attracted growing interest from scholars in the fields of environmental and energy economics. Adopting a range of conventional methodological approaches to time series and panel data analysis, numerous studies

[3]While there is an ongoing dispute over the optimal relative shares of renewable energy resources and nuclear power in electricity generation in France (Shirizadeh & Quirion, 2021), the share of electricity generation by nuclear power is to be reduced to 50% by 2035 as per government policy (World Nuclear Association, 2022).

have examined this relationship over different time periods and with different geographical scopes. As suggested by Sharif et al. (2020), existing research in this area can be broadly categorized into five groups according to their findings: studies that indicate a unidirectional causality relationship from renewable energy use to $CO_2$ emissions (Farhani & Shahbaz, 2014; Jaforullah & King, 2015; Özbuğday & Erbas, 2015; Apergis & Payne, 2015; Long et al., 2015; Bilgili et al., 2016; Bulut, 2017; Liu et al., 2017b; Khan et al., 2018; Salazar-Núñez et al., 2021); those arguing that $CO_2$ emissions influence the production and consumption of renewable energy (Sadorsky, 2009; Menyah & Wolde-Rufael, 2010; Shafiei & Salim, 2014; Leitão, 2014; Jebli & Youssef, 2015; Paramati et al., 2017); works suggesting a bidirectional causal association between $CO_2$ emissions and the production and consumption of renewable energy (Apergis et al., 2010; Dogan & Seker, 2016; Dong et al., 2017; Waheed et al., 2018); the ones that imply no causal link between $CO_2$ emissions and the production and consumption of renewable energy (Qi et al., 2014; Bento & Moutinho, 2016; Jebli et al., 2016; Saidi & Mbarek, 2016; Boontome et al., 2017; Jebli & Youssef, 2017; Liu et al., 2017a); and finally studies with mixed or indecisive results on this relationship (Zeb et al., 2014; Apergis & Payne, 2014; Sebri & Ben-Salha, 2014; Ang & Su, 2016; Bélaïd & Youssef, 2017; Sinha et al., 2018; Adams & Nsiah, 2019; Chen et al., 2019; Sharif et al., 2020; Rodrigues et al., 2020).

All the empirical studies listed above have based their analyses upon annual data, with the exception of the work of Sharif et al. (2020), which makes use of monthly data disaggregated from annual series. As emphasized by Adewuyi & Awodumi (2017), this body of literature remains inconclusive on the topic, both at single and multi-country levels. Interestingly enough, some studies with the same geographical scope (i.e. region) and overlapping study periods have yielded inconsistent results (see for example the works of Sebri & Ben-Salha (2014), Dong et al. (2017) and Liu et al. (2017b) on BRICS countries). This highlights the need for the recognition of the specificity of different research settings, and calls for further investigation into the matter, possibly by means of more advanced methodologies that are better fitted to answering the question on the nexus between renewable electricity production and $CO_2$ emissions.

From a methodological point of view, a potential drawback of most existing empirical studies in this area is the fact that they seek (positive or negative) causal links between $CO_2$ emissions and renewable energy production and consumption, while relying on methods and/or measures that are inherently inappropriate for drawing such causal inferences. For instance, commonly used emissions indicators (like the ones used in the present study; see Section 2.1) are typically estimated by considering the contribution of the combustion of carbon-based fuels (such as coal, gas and various fuel oil products) to $CO_2$ emissions, and on the assumption that renewable energy sources (also referred to as clean energy sources) or even biofuels[4] create little to no $CO_2$. Hence, the

[4]See for example the description of $CO_2$ emissions data from the BP Statistical Review of World Energy (`https://www.bp.com/en/global/corporate/energy-economics/statistica`

change in renewable energy production and consumption cannot necessarily be considered to be the "cause" of the change in the values of such emissions variables. Moreover, prior research has often used composite emissions indicators that may embrace, but are not necessarily limited to emissions from electricity production (see for example Apergis et al., 2010; Adams & Nsiah, 2019; Dogan & Seker, 2016). This variable choice could possibly lead to the omitted-variable bias in statistical models, and question the validity of any causal claim about the relationship between $CO_2$ emissions and renewable energy production and consumption.

With reference to the points raised above, it could be argued that most causal claims in this context can be considered indications of correlation, precedence (Leamer, 1985) or temporal relation (Granger & Newbold, 2014) based on a set of theoretical assumptions about the data. One way to overcome this problem of causal interpretation involves the use of advanced predictive modeling techniques to characterize the "predictive impact" (distinguished from causal impact) of renewable energy production on $CO_2$ emissions. Instead of attempting to make a theoretically unsupported causal claim, this alternative approach helps provide answers to questions of great practical importance such as (1) Given an increase or decrease in renewable energy production, what would be the predicted change in emissions in the electricity sector? and (2) Under a higher-renewable-electricity-production scenario, what would be the optimal share of each source in the the renewable energy mix for minimizing predicted emissions in the electricity sector? This proposed predictive framework derives its legitimacy from the fact that an increase in the share of the package of clean energy sources is expected to result in a decrease in the share of high-carbon energy sources as main drivers of $CO_2$ emissions.

On another note, while some existing works have considered the effect of specific types of renewable energy sources like hydroelectricity (Long et al., 2015; Khan et al., 2018) or wind and solar energies (Qi et al., 2014) on $CO_2$ emissions, there has been a general trend towards viewing renewable energy as a single variable and overlooking potential disparities between different types of renewable energy in terms of their impact on $CO_2$ emissions. More importantly, little attention has been paid to the extent to which different renewable energy sources can be influenced by climate change. In order to fill these gaps, the present study distinguishes between constituent sources of the renewable energy block, and focuses further on the role of "climate-related" renewable energy sources in reducing $CO_2$ emissions from electricity production. Consistent with the definition of Engeland et al. (2017), climate-related renewable energy (hereafter referred to as CRE) sources are represented in this paper by wind, solar photovoltaics, and small-scale run-of-river hydroelectricity energy sources. What justifies the appellation "climate-related renewable energy" is the fact that the availability and sporadicity of these resources are dependent on climate factors

l-review-of-world-energy/using-the-review/methodology.html.html#accordion_carbon) such as air temperature, wind speed, solar radiation, precipitation, and river runoff. Consequently, among different sources of energy, CRE sources are most affected by climate change.

The main purpose of this work is to characterize the predictive impact of CRE electricity production on $CO_2$ emissions in the French electric power system over the 2013-2021 period. To do so, a machine learning-based empirical modeling framework is employed to first evaluate the importance of different types of CRE electricity production in predicting $CO_2$ emissions, and specify the marginal effect of each CRE source on the predicted outcome of the model. Through a counterfactual analysis, the predictive impact of CRE production potential (as proxied by climate-derived energy indicators) on $CO_2$ emissions is then quantified. This analysis reveals if exploiting the full potential of CRE sources, which is equivalent to an increase in the share of CRE electricity production and a decrease in the share of non-CRE sources, would result in significantly lower predicted emissions over the study period. Finally, four counterfactual scenarios of increased CRE production over the study period are explored, and the optimal mix of CRE sources for minimizing predicted emissions under each scenario is identified. This analysis is complemented by the identification of the optimal CRE mix that would minimize the intermittency of CRE electricity production in France from 2013 to 2021. By limiting the scope of the study to $CO_2$ emissions from electricity production (instead of using composite or total emissions indicators), the present research undermines the possibility of the precedence of emissions over renewable electricity generation. This provides a sound conceptual basis for delineating the predictive effect of CRE electricity production on $CO_2$ emissions.

The contributions of this study to the literature on the relationship between carbon emissions and electricity production from renewable and non-renewable energy sources are manifold. First, while most of the existing studies in this area are based on annual or monthly data, the present research capitalizes on emissions and energy indicators data with a high (i.e. daily) temporal resolution. Indeed, in a similar context to that of the present study, using data with coarse temporal resolution (e.g. annual or monthly) leads to disregard of intra-monthly or intra-annual variability and intermittency of renewable energy sources that depend on climate (see Gernaat et al., 2021). Second, in the attempt to model the relationship between carbon emissions and electricity production from renewable and non-renewable energy sources, this paper considers all categories and subcategories of fuel types. This kind of fine-grained analysis has rarely been undertaken in the energy economics literature. Third, to the best of the author's knowledge, this study is the first to quantify the predictive impact of the (unexploited) CRE electricity production potential on energy-related $CO_2$ emissions. To this should be added the methodological contributions towards counterfactual estimation of $CO_2$ emissions based on realizable CRE electricity production, particularly in terms of the weighting scheme for energy indicators (see Section 2.2.2). Fourth, by decomposing the CRE package into its constituent elements (i.e. wind, solar photovoltaics, and run-of-river hydroelectricity) rather than viewing it as a unified block, the present study is able to determine the optimal share of individual sources within the package that minimizes counterfactual predictions of $CO_2$ emissions under near-feasible to idealistic hypothetical CRE production scenarios. This scenario-based approach has important implications for renewable energy development and management in France, since it provides evidence of the relative importance of each CRE source with regard to emissions reduction in the electricity sector. Finally, in the evaluation of the predictive impact of CRE electricity production on $CO_2$ emissions under the proposed scenarios, this research takes into account the intermittency of CRE sources that is motivated by the natural variability of climate factors. So far, this striking aspect of renewable electricity generation has been largely neglected in the studies on the dynamics between emissions and renewable energies.

The remainder of this paper is structured as follows. Section 2 describes data and methodology of the analysis. The results are presented in Section 3. A discussion of the study limitations and a few suggestions for future research are provided in Section 4. The paper concludes with a summary of the key findings and the empirical contributions made to the existing literature (Section 5).

## 2. Materials and Methods

### *2.1. Data*

#### *2.1.1. Realized $CO_2$ emissions and electricity production by different fuel types*

Consolidated and final half-hourly data on $CO_2$ equivalent emissions from electricity production (g/kWh), and the electrical power production by different fuel types (MW) in France from January 1, 2013 to August 31, 2021 were obtained from the éCO2mix data set, provided by the electricity transmission system operator of France and available on the Open Data Réseaux Énergies (ODRÉ) platform.[5] The emissions indicator represents $CO_2$ emissions released only by the consumption of primary fuel used in power plants, and is calculated based on the relative contribution of fuel oil, coal, gas and biofuel energy sources to $CO_2$ emissions. For the sake of this study and consistent with the nature of emissions data, cross-border physical power exchange (with England, Spain, Italy, Switzerland, Germany and Belgium) and the power consumed by pumps in pumped-storage hydroelectricity systems were disregarded. Furthermore, in light of methodological considerations, the analysis was restricted to non-negative values of power production, and negawatts (negative megawatts), if any, were set to zero. The original data were aggregated to average daily values,[6] and all power values in MW were converted to energy values in kWh. Table 1 presents the the summary statistics for the original emissions and en-

[5] `https://opendata.reseaux-energies.fr/`. This platform is a subdivision of the open platform for French public data (`https://www.data.gouv.fr/`)

[6] This aggregation is necessary to ensure consistency in sampling frequency across the different data sets used in this study.

ergy indicators used for empirical modeling (hereafter referred to as "realized" emissions and energy indicators).

Table 1: Summary statistics of realized daily emissions and energy indicators over the study period (from January 1, 2013 to August 31, 2021) based on the data provided by the electricity transmission system operator of France. $CO_2$ emissions values are expressed in grams (g). The measurement unit of the electric energy produced by different fuel types is kWh.

| Emissions/Energy Indicator | Mean | Max | Min | SD |
|---|---|---|---|---|
| $CO_2$ Emissions per kWh | 46.45 | 124.06 | 8.56 | 22.84 |
| Fuel Oil (Combustion Turbine) | 674,858 | 20,340,000 | 0 | 1,125,804 |
| Fuel Oil (Cogeneration) | 3,081,020 | 8,260,000 | 725,000 | 1,814,021 |
| Fuel Oil (Other) | 2,527,667 | 59,987,500 | 48,500 | 4,374,613 |
| Coal | 20,267,562 | 137,553,500 | 0 | 24,757,978 |
| Gas (Combustion Turbine) | 649,438 | 11,748,000 | 0 | 1,635,438 |
| Gas (Cogeneration) | 29,434,352 | 78,129,500 | 4,609,500 | 24,688,399 |
| Gas (Combined Cycle Turbine) | 47,794,491 | 138,265,500 | 0 | 40,036,372 |
| Gas (Other) | 1,906,760 | 14,620,500 | 312,000 | 1,687,226 |
| Nuclear | 1,057,000,000 | 1,458,000,000 | 530,400,000 | 166,199,782 |
| Wind | 70,972,076 | 321,939,000 | 3,475,500 | 54,964,062 |
| Solar Photovoltaics | 25,356,374 | 75,390,000 | 1,288,000 | 15,205,422 |
| Hydroelectricity (Run-of-river) | 113,731,842 | 190,018,500 | 34,160,000 | 35,640,184 |
| Hydroelectricity (Lake) | 45,365,080 | 117,566,500 | 6,776,500 | 19,626,978 |
| Hydroelectricity (Pumped-storage) | 16,000,279 | 41,715,000 | 1,621,500 | 7,322,262 |
| Biofuel (Waste) | 11,637,540 | 15,270,000 | 6,078,000 | 1,524,737 |
| Biofuel (Biomass) | 6,151,799 | 11,123,000 | 2,792,000 | 1,572,375 |
| Biofuel (Biogas) | 5,918,516 | 8,558,500 | 2,904,000 | 1,401,570 |

In addition to energy indicators, time-based features (i.e. year and month of the year) were created with integer encoding[7] and included in the empirical model as numerical control variables to account for possible year and month seasonality information in the data. The ultimate data set used for empirical modeling includes 3165 daily observations (from January 1, 2013 to August 31, 2021) with 19 independent variables (consisting of 17 energy indicators and 2 time-based features), and the natural logarithm[8] of $CO_2$ emissions per kWh of electricity generated as the response variable.

*2.1.2. CRE electricity production estimates derived from climate variables*

Climate variables such as air temperature, wind speed, solar radiation, precipitation, and river runoff can be transformed into potential renewable energy indicators by means of physical or statistical models or a combination of both. Numerical climate models for estimating CRE production potential are useful for delineating important climate-driven changes in the energy sector both in the short and the long term.[9]

[7]The use of integer encoding is permitted since the categories have a natural ordering.

[8]This transformation is necessary for empirical modeling purposes, i.e. to avoid potential negative predicted values of the response variable. Predicted emissions are back-transformed for the presentation and visualization of results.

[9]See Engeland et al. (2017) for a review of the foundations of such models, and a summary of the studies on the nexus between climate variability and renewable electricity production.

By Using a combination of physical and statistical models and considering the available installed energy capacity, the Copernicus Climate Change Service (C3S) at the European Centre for Medium-Range Weather Forecasts (ECMWF)[10] has provided a set of energy indicators for Europe derived from gridded reanalysis data on climate variables (Hersbach et al., 2020). This data set serves as a critical reference for evaluating the quality of climate-to-energy conversion models. The present study makes use of gridded and aggregated data over France on daily onshore wind, solar photovoltaics, and run-of-river hydroelectricity energy indicators from this collection (Ho et al., 2020; Saint-Drenan et al., 2018) for the period between January 1, 2013 and August 31, 2021.

The estimated energy indicators derived from climate variables are used as proxies for energy production "potential", and therefore referred to as "realizable" energy indicators further on in this paper. Indeed, they are assumed to represent the level of CRE electricity production that could be attained given the climate conditions of the study area (represented by the set of grid points within the area) over the period of interest.[11] Table 2 presents the the summary statistics for these energy indicators.

Table 2: Summary statistics of realizable daily climate-related renewable energy indicators over the study period (from January 1, 2013 to August 31, 2021) based on the data provided by the Copernicus climate change service (C3S). The realizable electric energy which could be generated by each fuel type is expressed in kWh.

| Energy Indicator | Mean | Max | Min | SD |
|---|---|---|---|---|
| Wind | 76,968,009 | 342,734,701 | 4,364,166 | 58,676,724 |
| Solar Photovoltaics | 29,833,555 | 58,314,090 | 1,865,264 | 13,638,494 |
| Hydroelectricity (Run-of-river) | 116,866,214 | 184,264,440 | 42,334,890 | 32,351,424 |

Figure 2 compares the distributions of original and estimated CRE indicators over the study period. A comparison of the respective medians in each panel of Figure 2 demonstrates a difference between the location of realized and realizable energy indicators. The significance of this location shift is further assessed using a non-parametric statistical test.

The results of the paired Wilcoxon signed-rank test (Wilcoxon, 1992; Conover, 1999) indicate that the median of the population of differences between estimated (model-derived) and original CRE indicators is greater than zero in all three cases ($p < 0.01$). From an empirical point of view, it thus seems perfectly legitimate to consider the so-called realizable energy indicators derived from climate variables as proxies for the CRE electricity production potential.

[10] `https://cds.climate.copernicus.eu/`

[11] This assumption, however, is subject to some limitations that are discussed in more detail in Section 4.

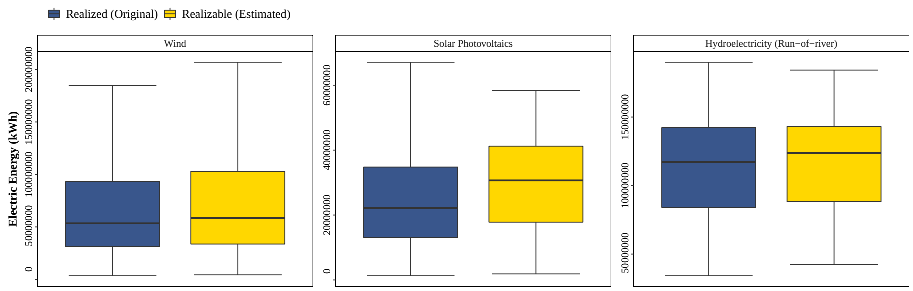


Figure 2: Compact box plots of realized (original) and realizable (estimated) CRE indicators over the study period. Note: The upper (lower) whisker extends from the hinge to the largest (smallest) value no further than 1.5 times the interquartile range. Data points beyond the whiskers are removed from the plot for the sake of better visualization.

### *2.2. Methodology*

#### *2.2.1. Empirical modeling of $CO_2$ emissions from electricity production based on various fuel types*

In order to empirically model (learn) the relationship between $CO_2$ emissions and the electric energy produced by different fuel types over the study period, stochastic Extreme Gradient Boosting (XGBoost) algorithm of Chen & Guestrin (2016) has been utilized.[12] XGBoost is a cutting-edge, speedy and highly performant decision-tree-based ensemble machine learning algorithm that can provide accurate predictions of a response variable by integrating the estimates obtained from a number of base models (trees). This predictive tool can model complex nonlinear relationships without assumptions about the data distribution, and is unsusceptible to multicollinearity. In its most general form, the tree ensemble model of the XGBoost algorithm that uses $K$ trained trees to predict the value of the response variable for a given data set with $N$ data points and $p$ features (predictors) $\{(x_i, y_i) \,|\, i = 1, ...N,\ x_i \in R^p,\ y_i \in R\}$ can be expressed as

[12]As with any other predictive model, XGBoost does not in itself imply "causal" relationships between variables. In this regard, by no means does the present modeling framework suggest that a change in low-carbon (in particular CRE) electricity production directly causes a change in $CO_2$ emissions. Indeed, in the calculation of emissions indicator in the éCO2mix data set, the contribution of low-carbon energy sources (nuclear, wind, solar photovoltaics, and hydroelectricity) has been considered equal to zero. Hence, such sources cannot directly "drive" emissions by definition. That being said, an increase in the share of nuclear and renewable energy sources inevitably translates into a decrease in the share of high-carbon energy sources as main drivers of $CO_2$ emissions. In this regard, a change in low-carbon electricity production is expected to be associated with a change in emissions. This is exactly where an empirical modeling framework like the one used here proves useful to characterize the "predictive" relationship between $CO_2$ emissions and electricity production from different fuel types (including low-carbon ones).

$$\hat{y}_i = \hat{f}(x_i) = \sum_{k=1}^{K} g_k(x_i) \quad g_k \in \mathcal{F} \tag{1}$$

where $\mathcal{F} = \{g(x) = w_{q(x)}\}\,(q : R^p \to J,\, w \in R^J)$ is the space of regression trees, $q$ is the structure of each individual (independent) tree that maps an observation to the corresponding leaf score $w$, and $J$ is the total number of leaves in the tree (Chen & Guestrin, 2016).

Each regression tree starts with a root node and is grown to a specific depth (i.e. the longest path from the root node to a leaf) by repeatedly splitting the training data based on all or some of the features in the feature space. This process results in a tree with a root node, a number of internal nodes (each of which split data points by one feature), and some leaves to which prediction scores (weights) are assigned. The ultimate predicted value of the response variable for a given observation is obtained by taking the sum of all the scores in the relevant leaves of individual trees. As proposed by Chen & Guestrin (2016), the choice of splitting points and the assignment of prediction scores in XGBoost are done by means of an improved and more regularized version of gradient boosting technique, in such a way as to minimize loss of an objective function that is composed of training loss and regularization (to avoid overfitting). Mathematically speaking, the tree building algorithm is reliant upon the minimization of

$$\mathcal{L} = \sum_{i=1}^{N} L(\hat{y}_i, y_i) + \sum_{k=1}^{K} \Omega(g_k) \tag{2}$$

where

$$\Omega(g_k) = \gamma J_k + \frac{1}{2}\lambda \sum_{j=1}^{J_k} w_{j,k}^2$$

Here, $L$ is a loss (cost) function (i.e. squared error, by default) and measures the difference between original values of the response variable $y_i$ and the predicted values $\hat{y}_i$ (Chen & Guestrin, 2016). $J_k$ and $w_{j,k}$ are the number of leaves and the prediction score assigned to the $j$-th leaf of the $k$-th regression tree, respectively. The parameter $\gamma$ is the minimum loss reduction required to further split the leaf node, and $\lambda$ is the L2 regularization on the prediction scores. These two, along with a number of tree-related parameters (together called hyperparameters of the model), cannot be estimated from data and need to be specified a priori.

In order to further minimize overfitting[13] and find the best model specifica-

[13]Although this study takes preventative measures to reduce overfitting, the reader's attention is drawn to the fact that overfitting should not, in principle, be cause for concern in the context of modern machine-learning models such as decision trees and ensemble meth-

tion, the present study combines extensive grid search hyperparameter tuning with repeated n-fold cross-validation ($n = 5$ with 5 repetitions). The metric used to evaluate the model performance for each hyperparameter configuration is the root-mean-square error (RMSE). There are 2592 variations in the hyperparameter search space and each variation is evaluated using repeated 5-fold cross validation with 5 repetitions, resulting in a total number of $2592 \times 25$ tree ensemble models to be trained and evaluated. The hyperparameter configuration that results in the minimum average RMSE across all folds is selected as the best tune. Possible hyperparameter values are determined mainly on the basis of recommendations of Boehmke & Greenwell (2019) and Thakur (2020). Table 3 presents the hyperparameter configurations used for evaluating tree ensemble models.

Table 3: Hyperparameter configurations used for evaluating tree ensemble models.

| Hyperparameter | Range | Default Value | Selected Values for Tuning |
|---|---|---|---|
| $\gamma$ | $[0,\infty)$ | 0 | {0.1,1,10} |
| $\eta$ | [0,1] | 0.3 | {0.05,0.1,0.2,0.3} |
| Maximum Depth | $\{1..\infty\}$ | 6 | {3..8} |
| Minimum Child Weight | $[0,\infty)$ | 1 | {7,10,20} |
| Column Sample by Tree | (0,1] | 1 | $\{\frac{6}{19},\frac{10}{19},\frac{14}{19}\}$ |
| Sub-sample | (0,1] | 1 | {0.3,0.5} |

Here, $\eta$ is the learning rate (also called shrinkage parameter), which shrinks prediction scores to prevent overfitting. $\gamma$ is the minimum loss reduction required to make a further split on a node of a given tree. Increasing $\gamma$ leads to a more conservative algorithm. The maximum depth parameter controls the number of terminal nodes in a a tree, and increasing its value makes the model more complex and more prone to overfitting. Minimum child weight determines the minimum sum of instance weight (hessian) required in a child node of a tree. A higher minimum child weight provides more conservative results. The column sample by tree parameter controls the fraction of columns (features) used for constructing each tree. Sub-sampling of columns takes place once for every tree constructed. Using values less than 1 for this parameter leads to a more conservative algorithm. The sub-sample parameter determines what fraction of data should be used to build trees in every boosting iteration. Using values less than 1 for this parameter leads to "stochastic" boosting, distinguished from "regular" boosting (which makes use of all points to grow a tree). In this analysis, the number of trees used for boosting is set to 50 and 100. In addition, the hyperparameter $\lambda$ is kept at the default value of 1.

It should be noted that the utilized algorithm is safeguarded against likely temporal autocorrelation in the data for two reasons. First, the data are divided, in a random manner, into training and validation data sets during the repeated 5-fold cross-validation process. Second, using stochastic boosting (as opposed to

ods. Belkin et al. (2019) show that modern (and complex) algorithms with near-perfect fit in training may still exhibit strong performance on unseen data.

regular boosting) makes the algorithm randomly select (without replacement) a proportion of the training data at each iteration. Therefore, the likelihood of neighboring observations being used by the algorithm at each iteration is very negligible.

Once the the model with the best tune is obtained from hyperparameter optimization, the "importance" of CRE sources in predicting emissions is calculated using the permutation feature importance algorithm with the RMSE ratio as the importance measure (Breiman, 2001; Fisher et al., 2019; Molnar, 2020), and 1000 repetitions. Permutation with repetition is performed with the aim of constructing the null distribution of importance measures. In its simplified form, the feature importance measure of a feature $p$ can be mathematically expressed as

$$\text{Feature Importance}_p = \frac{\text{RMSE}(y_i, \hat{f}(x_i^{perm:p}))}{\text{RMSE}(y_i, \hat{f}(x_i))} \quad (x_i \in R^p;\, y_i \in R) \tag{3}$$

where $x_i^{perm:p}$ is the $i^{\text{th}}$ instance with the $p^{\text{th}}$ feature replaced by a randomly sampled value, without replacement, from another instance. Being based on resampling without replacement, the permutation feature importance algorithm of Fisher et al. (2019) allows for conducting a permutation test with the null hypothesis that the importance of feature $p$ is 1:

$$H_0 : \text{Feature Importance}_p = 1 \tag{4}$$

If the $p^{\text{th}}$ feature is not important in the prediction, one should expect that the values for its feature importance measure be around 1. In point of fact, the proposed permutation test provides a framework for computing confidence intervals and p-values from resampling without replacement, and allows for determining statistical significance of a feature's importance.

As a complement to the features' importance evaluation, the influence of CRE electricity production on the prediction of the tree ensemble model is evaluated and visualized using mean-centered accumulated local effects (ALE) (Apley & Zhu, 2020). In machine learning, the ALE of a feature at a certain value is interpreted as the main effect of the feature at that value compared to the average prediction of the data (Molnar, 2020). By aggregating the calculated effects at different values, ALE plots–as unbiased alternatives to partial dependence plots–are hence able to show the (possibly nonlinear) relationship between the response variable and a given input feature. In mathematical terms, the mean-centered[14] ALE of a continuous feature $p$ at a given value $x$ is estimated as

$$\hat{f}_{p,ALE}(x) = \hat{\hat{f}}_{p,ALE}(x) - \frac{1}{N}\sum_{i=1}^{N} \hat{\hat{f}}_{p,ALE}(x_p^{(i)}) \tag{5}$$

[14]Mean-centering the ALE plot makes the average effect over the data be zero.

where

$$\hat{\tilde{f}}_{p,ALE}(x) = \sum_{l=1}^{l_p(x)} \frac{1}{n_p(l)} \sum_{i:x_p^{(i)} \in N_p(l)} \left[ \hat{f}(z_{l,p}, x_{\setminus p}^{(i)}) - \hat{f}(z_{l-1,p}, x_{\setminus p}^{(i)}) \right]$$

To estimate $\hat{\tilde{f}}_{p,ALE}(x)$, the distribution of the feature of interest $p$ is divided into a number of intervals (grids) denoted by $N_p(l)$, with $n_p(l)$ being the number of feature points that lie within the interval $N_p(l)$. The inner sum adds up the "effects" of all data points within such an interval (i.e. the differences in predictions, if the value of the feature of interest is replaced with the starting and end points of the given interval, namely $z_{l-1,p}$ and $z_{l,p}$). This sum is then divided by the number of feature points in this interval to obtain the average difference of the predictions for this interval. Finally, the outer sum accumulates the average effects across all intervals up to and including the interval $l_p(x)$ to which $x$ belongs (Apley & Zhu, 2020; Molnar, 2020). In order to define the aforementioned intervals, the present study makes use of the percentiles of the distribution of features. A distinct advantage of this choice is that each interval will contain the same number of data points. However, in this approach the length of intervals used for the calculation of ALE may not be the same.

*2.2.2. Counterfactual estimation of $CO_2$ emissions based on realizable CRE electricity production*

Once the predictive relationship between $CO_2$ emissions and electricity production from different fuel types is learned by the empirical model, original (realized) electricity production indicators are replaced with new conjectural values, and predictions of counterfactual emissions are generated from the model. The objective of this section is to describe a hypothetical yet achievable scenario, in which the maximum possible electric energy is generated from climate variables of the study area at a given time point, and the share of energy from non-CRE sources is reduced on a pro rata basis. Analyzing such a scenario is the first step towards quantifying the effect of increased share of CRE electricity production on predicted $CO_2$ emissions.

On a given day $t$, if the realizable value of a CRE source is greater than its corresponding realized value, the former is used as the new energy indicator and the difference between the realizable and realized values is regarded as the "unexploited" electricity production potential. Otherwise, the indicator is kept at the realized value, assuming that the potential for CRE electricity production has already been fully exploited.New counterfactual shares of CRE sources in the total electricity production are calculated based on the new energy indicators. In the next step, new counterfactual shares of non-CRE sources are calculated in such a way as to keep the total electricity production at the original level. This method of share reallocation guarantees a pro rata contribution of each non-CRE source to electricity production in the new setting. In other words, an increase in the share of the CRE package is counterbalanced by a proportional decrease in the share of each source in the non-CRE package.[15] In formal notation,

$$\text{Share}^{(\text{new})}_{i,t} = \text{Share}^{(\text{old})}_{i,t} \times \frac{1 - \sum_j \text{Share}^{(\text{new})}_{j,t}}{\sum_i \text{Share}^{(\text{old})}_{i,t}} \tag{6}$$

where

$$\begin{aligned}
\text{Share}^{(\text{old})}_{i,t} &= \frac{\text{Energy}^{(\text{old})}_{i,t}}{\sum_{\mathcal{S}} \text{Energy}_t}, \\
\text{Share}^{(\text{new})}_{j,t} &= \frac{\text{Energy}^{(\text{new})}_{j,t}}{\sum_{\mathcal{S}} \text{Energy}_t}, \\
\text{Energy}^{(\text{new})}_{j,t} &= \max\left(\text{Energy}^{(\text{realized})}_{j,t}, \text{Energy}^{(\text{realizable})}_{j,t}\right)
\end{aligned}$$

The superscripts (new) and (old) characterize new conjectural and original (realized) values, respectively. In addition, $i \in \mathcal{S} \setminus \{\text{W}, \text{PV}, \text{ROR}\}$, $j \in \{\text{W}, \text{PV}, \text{ROR}\}$, and $\mathcal{S}$ is the set of all energy indicators. W, PV and ROR denote wind, solar photovoltaics and run-of-river hydroelectricity, respectively. For non-CRE sources, new energy indicators are then calculated by multiplying their corresponding new share by the original sum of electricity production by different fuel types:

$$\text{Energy}^{(\text{new})}_{i,t} = \text{Share}^{(\text{new})}_{i,t} \times \sum_{\mathcal{S}} \text{Energy}_t \tag{7}$$

To test whether there is statistically significant difference between the locations of counterfactual and observed emissions in this scenario, the non-parametric Wilcoxon signed-rank test is used. The null hypothesis of this test is that the median difference between pairs of counterfactual and observed emissions is greater than or equal to zero. The rejection of the null hypothesis leads to the conclusion that, with the same level of total electricity production, increasing the share of energy from CRE sources (i.e. exploiting their full potential) decreases $CO_2$ emissions from electricity production.

### *2.2.3. Characterizing the optimal mix of CRE sources for minimizing predicted $CO_2$ emissions under different production scenarios*

In another attempt to specify the effect of CRE electricity production on predicted $CO_2$ emissions, different CRE production scenarios–from near-realistic to ambitious–are investigated and the optimal mix of CRE sources for minimizing

[15] While in this proposed framework the relative share of nuclear energy (as the main source of electricity in France) in the non-CRE package remains unaltered (with a still higher share of other sources in the non-CRE package), policy concerns are mainly included in the direct substitution of CRE sources for nuclear energy and the likely exclusion of nuclear power plants in the future.

emissions under each scenario is identified. Such scenarios would have been realized had the CRE electricity capacity been higher over the study period. Indeed, this analysis provides an overview of the relative importance of each CRE source (wind, solar photovoltaics and run-of-river hydroelectricity) within the CRE package with regard to the reduction of predicted $CO_2$ emissions.

With this aim and following the methodology of François et al. (2016), new CRE indicator daily series from Section 2.2.2 are normalized so that the mean production of each source equals the mean total daily electricity production over the study period (the same notation as above):

$$\text{Energy}_{j,t}^{(\text{normalized})} = \frac{\text{Energy}_{j,t}^{(\text{new})}}{\langle \text{Energy}_{j,t} \rangle} \times \langle \sum_{\mathcal{S}} \text{Energy}_t \rangle \tag{8}$$

where $\langle \rangle$ is the temporal mean operator. In this framework, a CRE mix electricity production can be described as a weighted sum of the three normalized CRE indicator series:

$$\text{Energy}_{CRE,t}(\zeta) = \zeta \sum_j \alpha_j \text{Energy}_{j,t}^{(\text{normalized})} \quad (\alpha_j \geq 0, \sum_j \alpha_j = 1) \tag{9}$$

where $\alpha_j$ is the share of the $j$-th CRE source in the CRE mix (package), and $\zeta$ is the ratio between the average energy produced by the CRE energy mix and the average total electricity production over the study period:

$$\zeta = \frac{\langle \text{Energy}_{CRE,t} \rangle}{\langle \sum_{\mathcal{S}} \text{Energy}_t \rangle} \tag{10}$$

If $\zeta = 1$, the mean daily CRE electricity production equals the mean daily total electricity production (i.e. the CRE electricity production is, on average, equal to the total electricity production over the study period). If $\zeta < 1$, a fraction of the total electricity production can, on average, be fulfilled by the CRE mix over the entire period. For the sake of this study, four hypothetical CRE electricity production scenarios with $\zeta = 0.25, 0.5, 0.75$ and 1 are considered.[16] For each scenario, all different combinations of wind, solar photovoltaics and run-of-river hydroelectricity in the CRE mix (with each share $\alpha_j$ ranging from 0 to 1 in increments of 0.05) are used to construct new CRE indicators. This results in 231 unique configurations of the CRE mix for each value of $\zeta$. Under a specific scenario and for each configuration, the shares of non-CRE sources are reallocated following the same approach as described in Section 2.2.2, and

[16] To be as realistic as possible, the present study sets the upper bound of $\zeta$ to 1. This means that scenarios with mean daily CRE electricity production exceeding the original mean daily total electricity production over the study period are not considered.

the empirical model introduced in Section 2.2.1 is utilized for generating predictions of counterfactual emissions. For a given scenario, the configuration that minimizes predicted mean daily $CO_2$ emissions over the study period is selected as the optimal CRE mix for emissions reduction.

Moreover, in order to account for the intermittency of CRE electricity production that is driven by the natural variability of climate factors (Seyedhashemi et al., 2021), the coefficient of variation (i.e. the ratio of the standard deviation to the mean) of the normalized daily CRE mix is calculated for each of the above-mentioned 231 configurations. The configuration that minimizes the coefficient of variation (hereafter denoted as CV) over the study period is selected as the optimal CRE mix for reducing intermittency. This additional analysis is offered to compare the optimal mix of CRE sources for minimizing $CO_2$ emissions under different production scenarios and the optimal mix of CRE sources for reducing problems of intermittency.[17]

All the analyses and data visualization in this study have been carried out in R software environment (R Core Team, 2020; Kuhn, 2008; Molnar et al., 2018; Hamilton & Ferry, 2018).

## 3. Results

### 3.1. *Influence of CRE electricity production on the prediction of $CO_2$ emissions*

Among all hyperparameter configurations for evaluating the tree ensemble models (see Section 2.2.1), the configuration with the hyperparameter values shown in Table 4 proved to minimize average RMSE across all folds (average RMSE= 0.062; average $R^2 = 0.986$), hence selected as the best tune.

Table 4: Optimal hyperparameter configuration of the tree ensemble model

| Hyperparameter | Best Tune |
|---|---|
| $\gamma$ | 0.1 |
| $\eta$ | 0.1 |
| Maximum Depth | 7 |
| Minimum Child Weight | 10 |
| Column Sample by Tree | $\frac{14}{19}$ |
| Sub-sample | 0.5 |
| Number of Trees | 100 |

The model with this optimal hyperparameter configuration was used as the base model for statistical analyses and prediction purposes of this study. The results of the permutation feature importance algorithm for the three CRE indicators based on the best empirical model are presented in Table 5. Here, the importance of each feature is measured by calculating the increase in the model's prediction error (in terms of the RMSE ratio) at each repetition, when

[17] The CV depends only on the shares of wind, solar photovoltaics and run-of-river hydroelectricity in the CRE mix, and is independent of the production scenario. Consequently, the optimal CRE mix for reducing intermittency is calculated once, regardless of the value of $\zeta$.

the values of that feature are shuffled (Molnar, 2020). A given CRE indicator is "important" for the prediction of emissions, if permuting its values increases the model RMSE (i.e. the model is reliant on the feature for the prediction). The CRE indicator is unimportant if permuting its values leaves the model RMSE unaltered (i.e. the feature is ignored by the model for the prediction).

Table 5: Permutation feature importance of CRE sources in predicting $CO_2$ emissions (number of repetitions = 1000).

| Feature | Importance (RMSE Ratio) | | |
|---|---|---|---|
| | 5th Percentile | Median | 95th Percentile |
| Wind | 1.171104 | 1.181664 | 1.192710 |
| Solar Photovoltaics | 1.007700 | 1.009709 | 1.011769 |
| Hydroelectricity (Run-of-river) | 1.252940 | 1.268573 | 1.282155 |

For all CRE indicators, the value 1 is outside the 90% confidence interval for feature importance estimates. This concludes that wind, solar photovoltaics and run-of-river hydroelectricity features are all important for the prediction of emissions resulting from the generation of electrical power at the 0.1 significance level (see Equation 4). Among the three CRE indicators, run-of-river hydroelectricity proves to be the most important feature for predicting emissions, followed by wind energy. The importance of solar photovoltaics feature is only marginal, since the 5th, 50th and 95th percentiles of the distribution of RMSE ratio for this feature are close to 1. This finding acts as an early indicator of the level of importance of each CRE source for emissions reduction in France.

The influence of each CRE source on the prediction of emissions by the best model is further assessed by calculating ALE values. Figure 3 presents ALE plots of wind, solar photovoltaics and run-of-river hydroelectricity features. In basic terms, multiple panels of this figure illustrate how the model predictions change–compared to the average prediction of the data–for different values of each CRE indicator.

As a first observation, the ALE curve is monotonically non-increasing for all CRE indicators.[18] This means that the prediction decreases or remains constant, compared to the average prediction, with increasing CRE electricity production–a finding consistent with expectations that renewable energy sources can contribute to the reduction of carbon emissions.[19] For wind, solar

[18]Although interval-wise effects are accumulated to construct a smooth ALE curve, the effects are estimated locally using different data points. Therefore, one should be cautious when interpreting the effect across intervals (Molnar, 2020).

[19]Special caution must be taken when evaluating and interpreting the predictive impact of individual low-carbon power generation technologies on $CO_2$ emissions in the electricity sector. A climate-driven increase (decrease) in the share of a given CRE source may not necessarily be counterbalanced by a decrease (increase) in the share of high-carbon energy sources (coal, fuel oil, gas and biofuel). For instance, an increase (decrease) in the share of solar photovoltaics due to greater (smaller) availability of solar resources may be offset by a decrease (increase) in the share of other non-polluting energy sources (e.g. nuclear, hydroelectricity, etc.) and not necessarily fossil-fuel fired electricity generation, hence leaving emission levels unaltered. A full discussion of the elasticity of inter-fuel substitution between different CRE sources,

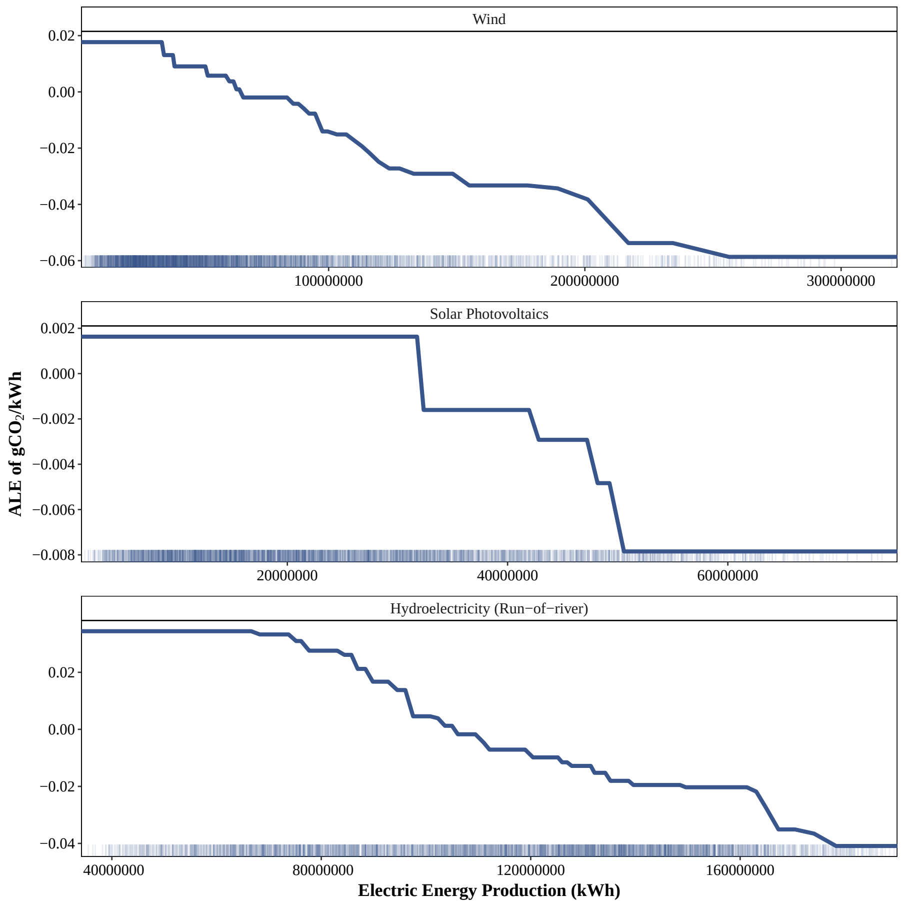


Figure 3: Mean-centered accumulated local effects (ALE) of CRE sources in predicting $CO_2$ emissions over the study period. The distribution of data points for each feature is displayed on the margin of horizontal axis.

photovoltaics and run-of-river hydroelectricity features, the prediction of $CO_2$ emissions remains approximately constant in intervals below the 30th, 70th and 13th percentile (corresponding to 34898000, 31780500, and 66610500 kWh of electrical energy produced), and above 99th, 92nd and 98th percentile (corresponding to 256266500, 50565500, and 178323000 kWh of electrical energy produced), respectively. Compared to the case of wind and run-of-river hydroelectricity features, the ALE function of the solar photovoltaics feature is constant on larger intervals. Furthermore, the range of change in the ALE of

between nuclear and CRE sources, and between different fossil fuels and CRE sources lies beyond the scope of this study.

solar photovoltaics (i.e. as the feature increases from the its minimum value to the maximum value) is relatively smaller than those of wind and run-of-river hydroelectricity features (0.009 for solar photovoltaics, compared to 0.076 for wind and 0.075 for run-of-river hydroelectricity). This is another important finding in the understanding of the significance of each CRE source for reducing emissions that are associated with the generation of electrical power.

### 3.2. *Estimated counterfactual $CO_2$ emissions based on realizable CRE electricity production*

Using new conjectural energy indicators constructed from realizable CRE electricity production (see Section 2.2.2), the best model can generate counterfactual estimates of $CO_2$ emissions, i.e. emissions that would have been realized provided that the full potential of CRE sources for electricity production had been exploited. Figure 4 compares observed (realized) and estimated counterfactual daily emissions based on new energy indicators in France from January 1, 2013 to August 31, 2021.

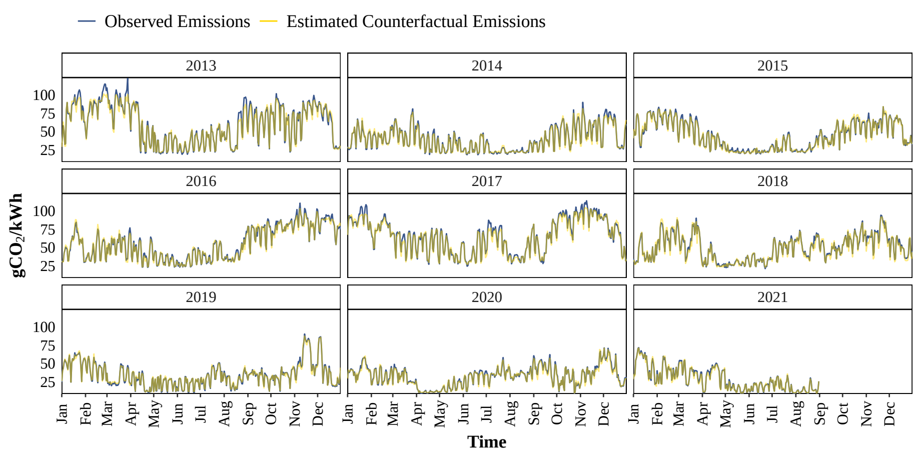


Figure 4: Observed (realized) and estimated counterfactual daily emissions based on realizable CRE electricity production over the study period. For visualization purposes, data for each year are presented in a separate panel.

The percentage difference between estimated counterfactual and observed daily emissions (calculated as [estimated counterfactual emissions$-$observed emissions)/observed emissions] $\times 100$) ranges from $-20.15$ to 19.91, with the average percentage difference being $-1.13$. This result highlights that exploiting the full potential of climate variables for electricity production would have, on average, resulted in 1.13% reduction in predicted $CO_2$ emissions in France over the study period. The distribution of the difference between estimated counterfactual and observed emissions over the entire study period is illustrated in Figure 5.

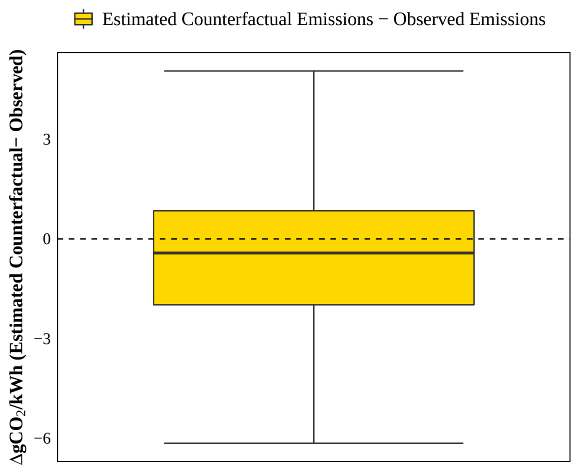


Figure 5: Box plot of the difference between estimated counterfactual emissions (based on realizable CRE electricity production) and observed emissions over the study period. Note: The upper (lower) whisker extends from the hinge to the largest (smallest) value no further than 1.5 times the interquartile range. Data points beyond the whiskers are removed from the plot for the sake of better visualization.

The null hypothesis of the non-parametric Wilcoxon signed-rank test (i.e. the median difference between pairs of counterfactual and observed emissions is greater than or equal to zero) is rejected at the 0.01 significance level, indicating that an increase in the share of energy from CRE sources (under a scenario where the maximum possible electric energy is generated from climate variables) would have been associated with a statistically significant decrease in $CO_2$ emissions from electricity production over the study period.

### *3.3. Optimal mix of CRE sources for minimizing predicted $CO_2$ emissions under different production scenarios*

This section presents the results of the analysis of the four electricity production scenarios outlined in Section 2.2.3. Under each scenario, counterfactual estimates of $CO_2$ emissions (generated by the base empirical model) are compared with observed emissions.

The first scenario corresponds to the situation where the mean daily CRE electricity production equals 25% of the mean electricity production over the study period ($\zeta = 0.25$). The empirical $\zeta$ based on the realized and realizable CRE electricity production (as defined in Sections 2.1.1 and 2.1.2) over the study period is 0.144 and 0.158, respectively. In other words, the CRE electricity production satisfied, on average, 14.4% of the total electricity production over the study period in reality. Had the full potential of CRE sources been exploited over the same period, this ratio would have risen to 15.8%. From these observations, it is clear that the first scenario is the most realistic and feasible among the four proposed scenarios. Figure 6 compares observed (realized) and estimated counterfactual daily emissions under the first scenario from January 1, 2013 to August 31, 2021.

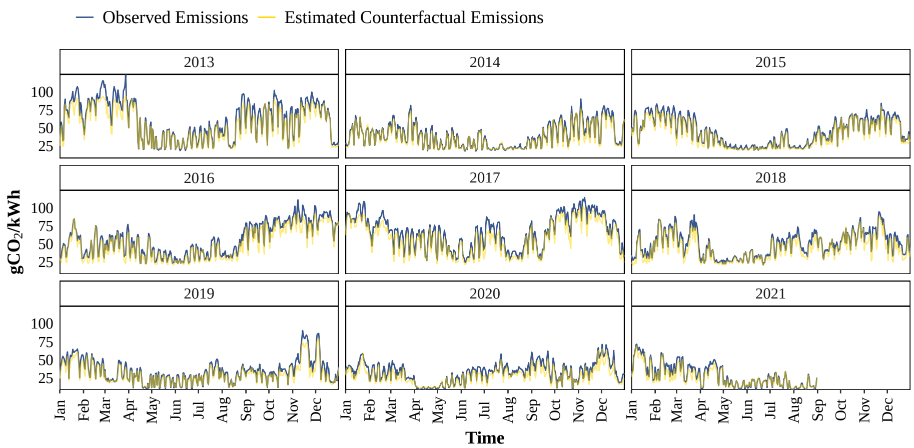


Figure 6: Observed (realized) and estimated counterfactual daily emissions over the study period under the first production scenario ($\zeta = 0.25$). For visualization purposes, data for each year are presented in a separate panel.

The percentage difference between estimated counterfactual and observed daily emissions for $\zeta = 0.25$ ranges from $-36.07$ to 35.69, with the average percentage difference being $-9.56$. This suggests that, under the first hypothetical CRE electricity production scenario, daily $CO_2$ emissions would have, on average, decreased by 9.56%.

The second scenario corresponds to the situation where the mean daily CRE electricity production equals 50% of the mean electricity production over the study period ($\zeta = 0.5$). Figure 7 compares observed (realized) and estimated counterfactual daily emissions under the second scenario from January 1, 2013 to August 31, 2021.

The percentage difference between estimated counterfactual and observed daily emissions for $\zeta = 0.5$ ranges from $-78.34$ to 36.25, with the average percentage difference being $-24.93$. This suggests that, under the second hypothetical CRE electricity production scenario, daily $CO_2$ emissions would have, on average, decreased by 24.93%.

The third scenario corresponds to the situation where the mean daily CRE electricity production equals 75% of the mean electricity production over the study period ($\zeta = 0.75$). Figure 8 compares observed (realized) and estimated counterfactual daily emissions under the third scenario from January 1, 2013 to August 31, 2021.

The percentage difference between estimated counterfactual and observed daily emissions for $\zeta = 0.75$ ranges from $-83.03$ to 32.32, with the average percentage difference being $-36.92$. This suggests that, under the second hypothetical CRE electricity production scenario, daily $CO_2$ emissions would have,

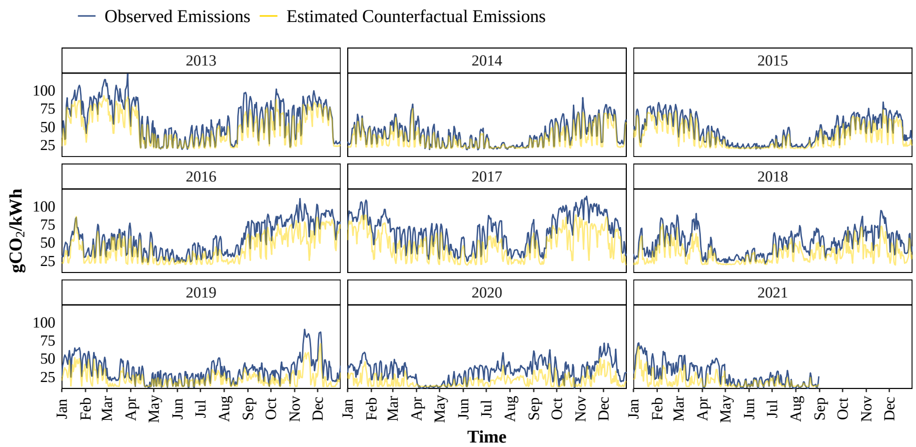


Figure 7: Observed (realized) and estimated counterfactual daily emissions over the study period under the second production scenario ($\zeta = 0.5$). For visualization purposes, data for each year are presented in a separate panel.

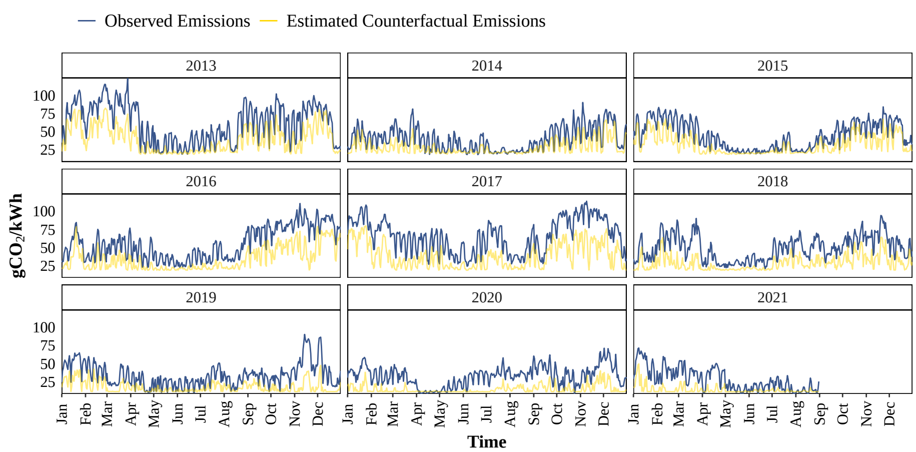


Figure 8: Observed (realized) and estimated counterfactual daily emissions over the study period under the third production scenario ($\zeta = 0.75$). For visualization purposes, data for each year are presented in a separate panel.

on average, decreased by 36.92%.

The fourth and last scenario corresponds to the situation where the mean daily CRE electricity production equals the mean daily total electricity production ($\zeta = 1$). This scenario is the most ambitious among the four proposed scenarios. Figure 9 compares observed (realized) and estimated counterfactual daily emissions under the fourth scenario from January 1, 2013 to August 31, 2021.

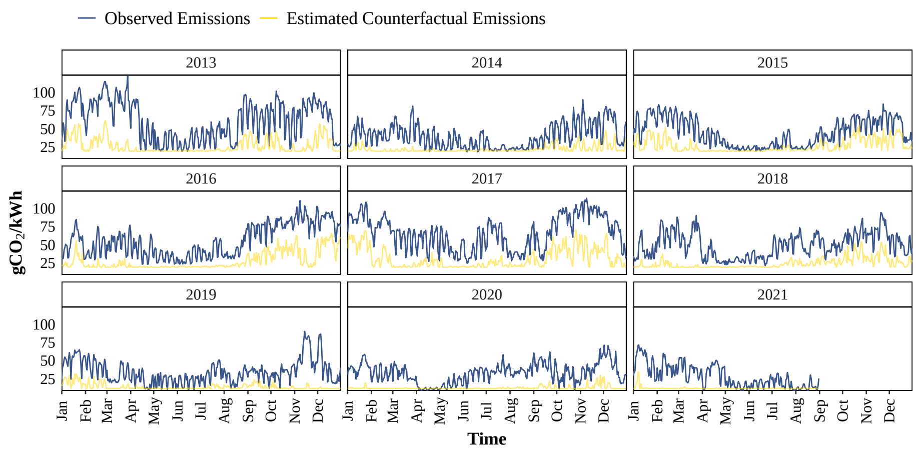


Figure 9: Observed (realized) and estimated counterfactual daily emissions over the study period under the fourth production scenario ($\zeta = 1$). For visualization purposes, data for each year are presented in a separate panel.

The percentage difference between estimated counterfactual and observed daily emissions for $\zeta = 1$ ranges from $-87.06$ to 32.32, with the average percentage difference being $-47$. This suggests that, under the fourth hypothetical CRE electricity production scenario, daily $CO_2$ emissions would have, on average, decreased by 47%.

The distributions of the difference between estimated counterfactual and observed emissions under the four proposed scenarios over the entire study period are illustrated in Figure 10.

The null hypothesis of the non-parametric Wilcoxon signed-rank test (i.e. the median difference between pairs of counterfactual and observed emissions is greater than or equal to zero) is rejected at the 0.01 significance level for all the proposed scenarios. The results confirm that an increase in the ratio of mean daily CRE electricity production to mean total electricity production, would have been associated with a statistically significant decrease in predicted $CO_2$ emissions in France from 2013 to 2021. From a practical and economic standpoint, this means that increasing the share of CRE electricity production and decreasing the share of non-CRE sources, while retaining the same level of total production, would have significantly reduced emissions over the study period.

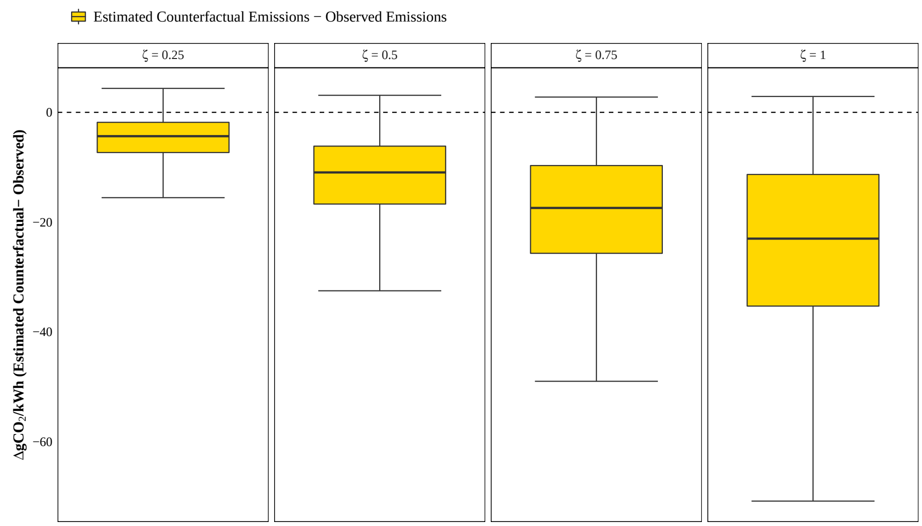


Figure 10: Box plot of the difference between estimated counterfactual emissions under different production scenarios ($\zeta = 0.25, 0.5, 0.75$ and 1), and observed emissions over the study period. Note: The upper (lower) whisker extends from the hinge to the largest (smallest) value no further than 1.5 times the interquartile range. Data points beyond the whiskers are removed from the plot for the sake of better visualization.

In the next step and in order to characterize the relative importance of each CRE source for reducing predicted $CO_2$ emissions from electricity production, the optimal CRE mix for emissions reduction under each proposed scenario is identified. Figure 11 depicts the predicted mean daily $CO_2$ emissions over the study period in the CRE mix space. The share of each source in the CRE package ranges from 0 to 1 in increments of 0.05, with the sum of shares being equal to 1.

Table 6 presents the optimal CRE mix for emissions reduction under the four proposed scenarios ($\zeta = 0.25, 0.5, 0.75$ and 1), and the corresponding predicted mean daily $CO_2$ emissions for each mix. These results offer additional evidence that run-of-river hydroelectricity is the most important source for reducing energy-related $CO_2$ emissions in France, with up to 75% share in the optimal CRE mix. The only exception is when $\zeta = 0.5$, where the share of run-of-river hydroelectricity is less than the share of wind in the optimal CRE mix. Solar photovoltaics has the smallest share of CRE sources in the optimal mix under all proposed scenarios.

These results should be considered in comparison with realized and realizable mean shares of the three sources in the CRE mix. The empirical mean shares of wind, solar photovoltaics and run-of-river hydroelectricity in the CRE mix based on the realized electricity production (as defined in Section 2.1.1) are 31.11%, 13.06% and 55.82%, respectively, with the corresponding mean daily $CO_2$ emissions being 46.44 $gCO_2/kWh$. In a similar vein, the empirical mean

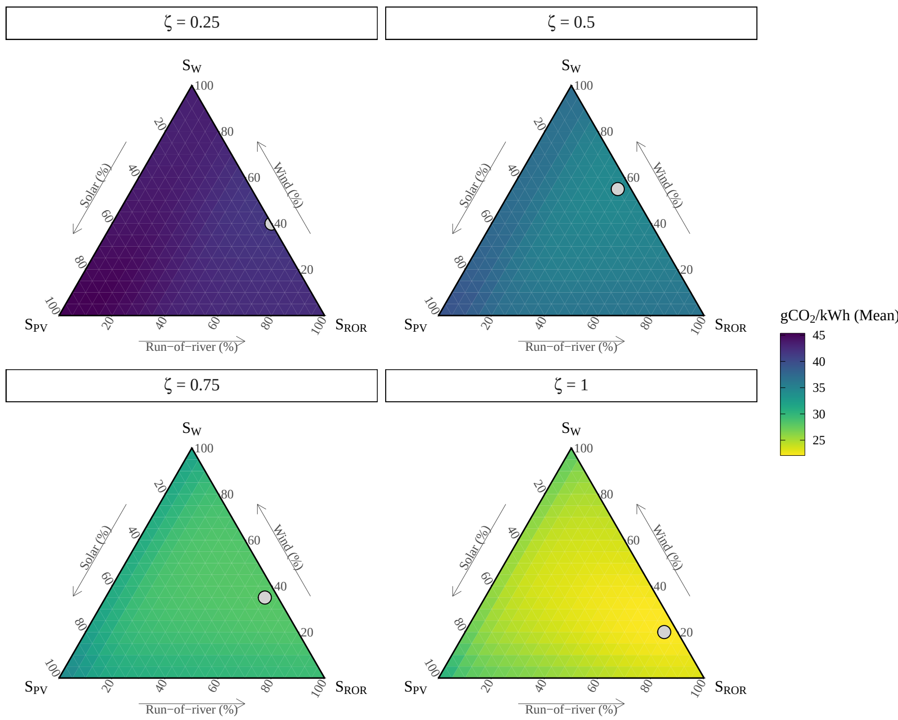


Figure 11: Ternary graphs of the mean daily $CO_2$ emissions over the study period for different shares of wind, solar photovoltaics and run-of-river hydroelectricity in the CRE mix under different production scenarios ($\zeta = 0.25, 0.5, 0.75$ and 1). The optimal CRE mix for emissions reduction is marked with a circle on each graph.

Table 6: Optimal CRE mix for emissions reduction under the four proposed scenarios, and the corresponding predicted mean daily $CO_2$ emissions.

| Scenario | Optimal CRE Mix | | | $gCO_2$/kWh (Mean) |
|---|---|---|---|---|
| | $S_W$ | $S_{PV}$ | $S_{ROR}$ | |
| $\zeta = 0.25$ | 40% | 0% | 60% | 41.56 |
| $\zeta = 0.5$ | 55% | 5% | 40% | 34.47 |
| $\zeta = 0.75$ | 35% | 5% | 60% | 28.15 |
| $\zeta = 1$ | 20% | 5% | 75% | 22.10 |

shares of wind, solar photovoltaics and run-of-river hydroelectricity in the CRE mix based on the realizable electricity production (as defined in section 2.1.2) are 31.84%, 14.98% and 53.17%, respectively, with the corresponding predicted mean daily $CO_2$ emissions being 45.7 $gCO_2$/kWh. From this comparison, two key findings emerge: (1) an increase in the share of CRE electricity production is associated with a decrease in average emissions from electricity generation, and (2) in both real and hypothetical contexts, the shares of different sources are not uniformly distributed within the CRE package.

To characterize the relative importance of each source for reducing the intermittency of CRE electricity production, the optimal CRE mix that minimizes CV over the study period is identified. This allows for comparison between the optimal CRE mix for emissions reduction under different scenarios and the optimal CRE mix for reducing intermittency. Indeed, the configurations that result in the minimum mean daily $CO_2$ emissions under the proposed scenarios do not necessarily result in low intermittency and high reliability of CRE electricity production over the study period. Figure 12 depicts the CV in the CRE mix space based on the normalized CRE indicator series over the study period. The share of each source in the CRE package ranges from 0 to 1 in increments of 0.05, with the sum of shares being equal to 1.

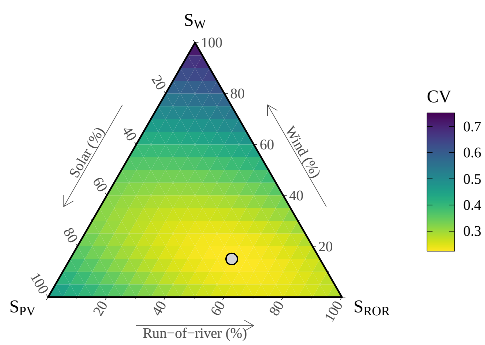


Figure 12: The coefficient of variation (CV) of the normalized daily CRE mix over the study period. The optimal CRE mix reducing intermittency is marked with a circle on the graph.

The shares of wind, solar photovoltaics, and run-of-river hydroelectricity in the optimal CRE mix for reducing the intermittency are 15%, 30% and 55%, respectively, with the corresponding CV being 0.22.[20] This CRE mix would result in the mean daily $CO_2$ emissions of 42.47, 35.5, 28.96 and 23 $gCO_2$/kWh under

[20]The empirical CV of the CRE package based on the realized and realizable CRE electricity production (as defined in Sections 2.1.1 and 2.1.2) are 0.31 and 0.30, respectively.

the first, second, third and fourth proposed scenario, respectively. Comparing these values with the optimal values in Table 6, it can be deduced that the optimal CRE mix for reducing intermittency is associated with slightly larger emissions than those of the optimal CRE mix for emissions reduction under the proposed scenarios.

These findings are important for two reasons. First, while the share of wind is larger than that of solar photovoltaics in the optimal CRE mix for emissions reduction under all the proposed scenarios, solar photovoltaics proves to have a larger share than wind in the optimal CRE mix for reducing intermittency. Second, among the three sources, run-of-river hydroelectricity has the largest share in the optimal CRE mix as it was the case for the optimal CRE mix for emissions reduction under three (out of four) proposed scenarios. The results of the present study can complement those of François et al. (2016), who determined the optimal CRE mix for maximizing energy penetration in France over the $1980-2012$ period ($S_W = 15\%$, $S_{PV} = 45\%$, $S_{ROR} = 40\%$, with the corresponding CV being 0.28). While beyond the scope of this paper, disentangling the complexities in finding the optimal balance among emissions reduction, intermittency reduction and energy penetration maximization might prove an important area for further investigation.

## 4. Discussion

On the part of European countries, the development of electricity generation from renewable sources and the European Union Emissions Trading Scheme (EU ETS)–the mainspring of the European Union's policy to reduce $CO_2$ emissions– are two important tools to address climate change. Nevertheless, complex interactions between these instruments, and in particular, the potential dampening effects of renewable electricity growth on emissions allowances prices, have raised considerable doubts on the feasibility of combining different targets and policies to effectively reduce carbon emissions (see del Río (2017) and Möst & Fichtner (2010)). A multidisciplinary economic analysis of this interaction by del Río (2017) indicates that most of the concerns on this matter are not supported by economic theory, and that the combination of the EU ETS and renewable energy-based electricity development should be favored. Notwithstanding, after several years of research, there is still no firm consensus on the environmental viability of combining emissions trading systems and renewable energy expansion. According to a recent theoretical study, in the long run, emissions trading systems may impede the expansion of renewable energy capacity rather than promoting it (Bersani et al., 2022). In this regard, a venue for future research includes (1) replicating the findings of the present research in other Member States of the EU ETS, and (2) empirically investigating the interaction between the EU ETS and the development of different types of CRE electricity production in Europe, especially by focusing on the economic value of renewable energy-induced emissions reduction considering the EU ETS allowance prices and potential carbon leakage.[21]

As with the majority of studies, the findings of this research have to be considered in the light of some limitations. The first limitation concerns the choice of proxies for CRE electricity production potential. It could be argued that in the process of estimating counterfactual $CO_2$ emissions, the construction of new CRE indicators relies only on the realizable CRE indicators that are greater than their corresponding realized values (see Section 2.2.2). In statistical terms, only "overestimated" CRE electricity production derived from climate variables is retained and "underestimated" values are disregarded. The assumption here is that, if the realized energy indicators are greater than the indicators estimated by the climate-to-energy model, the potential for CRE electricity production has already been fully exploited and there is therefore no point in utilizing the climate-derived energy indicators in such a case. In justification of this assumption, an argument can be made that in 63.06%, 68.53% and 62.27% of daily observations, the so-called realizable indicators are greater than their corresponding realized values for wind, solar photovoltaics, and run-of-river hydroelectricity, respectively. Moreover, for wind, solar photovoltaics, and run-of-river hydroelectricity, the normalized root-mean-square deviation (defined as the square root of the quadratic mean of the differences between realizable and realized values, divided by the range of realized values) that is associated with "overestimation" instances is, respectively, 1.6, 1.45 and 1.32 times greater than the normalized root-mean-square deviation associated with "underestimation" instances for the same energy source. Unless the models behind the data described in Section 2.1.2 are prone to systematic overestimation, one possible conclusion that can be drawn from these results is that the transformation of gridded climate variables into energy indicators would provide more of a reference to assess the "potential" of CRE sources than an estimation of the level of energy generation practically achieved. This assumption is the primary limitation to the interpretation of the results presented in Section 3.2. Nevertheless, in the absence of better alternatives, climate-to-energy conversion models are the only tools available to measure the (unexploited) potential for electricity production from CRE sources.

Another limitation of the present study includes the non-consideration of within-country (i.e. regional) dynamics between carbon emissions and electricity production from renewable energy sources. This limitation is mainly rooted in the lack of data at the regional level on $CO_2$ emissions from electricity production at daily time scale. Upon availability of more spatially-fine-grained data on emissions and climate-derived energy indicators, future studies could explore such regional dynamics. To provide a starting point for discussion and further research, Figure 13 illustrates the average share of daily electricity production from wind, solar and hydroelectric sources (run-of-river, lake and pumped-storage) by administrative region in metropolitan France over the study period. As shown

[21] As a relevant work in this area, see the study of Beltrami et al. (2021) that has examined the value of carbon emission reduction induced by renewable energy production in Italy.

in this figure, the share of wind and solar photovoltaics energy in electricity production is higher in northern and southern (coastal) regions, respectively. As expected, hydropower generation is more pronounced in mountainous regions, with the Auvergne-Rhône-Alpes region having a high average share of 44.55% in total electricity production in France over the study period.

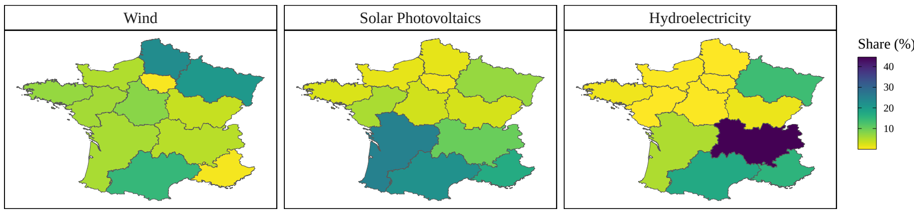


Figure 13: Average share of daily electricity production from wind, solar and hydroelectric sources (run-of-river, lake and pumped-storage) by region in metropolitan France over the study period.

Additionally, neither energy storage and residual load variability nor the economic and social cost of increasing the share of renewable energy vis-à-vis emissions reduction is considered in this paper. The study of Shirizadeh & Quirion (2021) has evaluated the relative contribution of renewable (wind, solar photovoltaics, run-of-river and lake hydroelectric) energy production, nuclear power and CCS technologies to the cost-optimal electricity mix in France, taking into account the social cost of carbon. The authors have found that a cost-optimal power mix consists of approximately 75% electricity production by renewable energy sources, and the remaining 25% is shared among nuclear power and fossil fuels, with or without CCS technologies. In a subsequent study, Shirizadeh et al. (2022) have examined the robustness of a renewable power system for France to key technology cost uncertainties by considering several cost scenarios. They have found that, although the cost-optimal electricity mix in France heavily depends on assumptions about technology costs, investments in the development of renewable energy should be prioritized even if those underlying cost assumptions prove to be wrong.[22] These findings, together with the results of the present study, may be important for policy and subsequent research.

Last but not least, it is worth noting that the principal outcomes of this explorative study are based on historical data and a counterfactual analysis of the nexus between CRE electricity production and $CO_2$ emissions in France. Hence, one should be careful when interpreting or extrapolating these results to other settings and/or time periods. For instance, in the majority of proposed scenarios, run-of-river hydroelectricity is identified as the CRE source of greater relative importance for reducing counterfactual predictions of $CO_2$ emissions

[22] Also see the study of Chu & Hawkes (2020) that has proposed a multi-objective optimization model to find the optimal mix of CRE sources in global electricity systems, considering cost, residual load variability, and portfolio output variability.

over the 2013-2021 period. However, the soundness of increased investment in the development of run-of-river hydroelectric power facilities (either by increasing the number of run-of-river power plants or improving the efficiency of installed power plants) to cut $CO_2$ emissions in the future is subject to the extent that France's run-of-river hydropower potential is affected by climate change, on top of cost considerations. A study based on the climate change projections and a set of scenario assumptions for future water use in Europe (Lehner et al., 2005) confirms that run-of-river hydropower potential (as measured by river discharges) will remain rather stable in the case of France for the time slices for the 2020s and 2070s. With the help of newly-developed climate projections, further research on this direction is warranted.

## 5. Conclusion

As a path forward to combat global climate change, the development of renewable energy share of electricity production remains the cornerstone of $CO_2$ emissions reduction in the electric power sector. Considering the dependence of the availability and sporadicity of climate-related renewable energy (CRE) sources (wind, solar photovoltaics, and small-scale run-of-river hydroelectricity) on climate factors, the relationship between $CO_2$ emissions and electricity production from these sources merits careful analysis. By means of a cutting-edge decision-tree-based modeling technique, this study characterized the relationship between daily CRE electricity production and energy-related $CO_2$ emissions in France and offered a framework for counterfactual analysis of such relationship over the 2013-2021 period.

The empirical analysis was undertaken in three steps. In the first step, the importance of CRE electricity production in predicting $CO_2$ emissions was assessed by means of the permutation feature importance algorithm. Furthermore, the nonlinear relationship between realized electricity production from CRE sources and predicted emissions was identified through accumulated local effects (ALE) plots. From the results, run-of-river hydroelectricity proved to be the most important feature among the three CRE sources for predicting emissions, followed by wind energy. Solar photovoltaics was shown to be of marginal importance in respect of predicting emissions. Next, the predictive impact of CRE electricity production potential (as proxied by climate-derived energy indicators) on $CO_2$ emissions was quantified. The results demonstrated that an increase in the share of energy from CRE sources–under a scenario where the maximum possible electric energy is generated from climate variables–would have been associated with a statistically significant decrease in $CO_2$ emissions from electricity production over the study period. Ultimately, four hypothetical CRE production scenarios were considered and the optimal mix of CRE sources for minimizing emissions under each scenario was determined. The findings confirmed greater relative importance of run-of-river hydroelectricity and wind energy within the CRE package with regard to the reduction of predicted $CO_2$ emissions. This step was complemented by the identification of optimal

CRE mix for reducing intermittency of CRE electricity production. This complementary analysis found further evidence for a higher share of run-of-river hydroelectricity in the CRE mix.

The findings of this research, while exploratory, can have important implications for renewable energy development and management in France, since they provide some support for the conceptual premise that replacing carbon-intensive energy sources with renewable ones reduces carbon emissions from electricity generation. Additionally, the findings cast a new light on the relative importance of each CRE source with regard to emissions and intermittency reduction in the electricity sector. Together these results might prove enlightening for policymakers who decide which renewable energy infrastructure investments should be given priority.

## 6. Acknowledgements

This work was realized in the course of a doctoral project at the University of Lille, funded by IÉSEG School of Management. The constant support of the doctoral advisor, Dr. Paolo Mazza, and financial support from the Institut Europlace de Finance (IEF) are gratefully acknowledged.

## References

Adams, S., & Nsiah, C. (2019). Reducing carbon dioxide emissions; Does renewable energy matter? *Science of the Total Environment*, *693*, 133288. doi:https://doi.org/10.1016/j.scitotenv.2019.07.094.

ADEME (2015). Un mix électrique 100% renouvelable ? Analyses et optimisations. URL: https://www.ademe.fr/sites/default/files/assets/documents/mix-electrique-rapport-2015.pdf accessed on 14 February 2022.

ADEME (2018). Trajectoires d'évolution du mix électrique 2020-2060. URL: https://librairie.ademe.fr/cadic/1176/evolution-mix-electrique-horizon-2020-2060-010655.pdf?modal=false accessed on 14 February 2022.

Adewuyi, A. O., & Awodumi, O. B. (2017). Renewable and non-renewable energy-growth-emissions linkages: Review of emerging trends with policy implications. *Renewable and Sustainable Energy Reviews*, *69*, 275–291. doi:https://doi.org/10.1016/j.rser.2016.11.178.

Ang, B. W., & Su, B. (2016). Carbon emission intensity in electricity production: A global analysis. *Energy Policy*, *94*, 56–63. doi:https://doi.org/10.1016/j.enpol.2016.03.038.

Apergis, N., & Payne, J. E. (2014). Renewable energy, output, CO2 emissions, and fossil fuel prices in Central America: Evidence from a nonlinear panel smooth transition vector error correction model. *Energy economics*, *42*, 226–232. doi:https://doi.org/10.1016/j.eneco.2014.01.003.

Apergis, N., & Payne, J. E. (2015). Renewable energy, output, carbon dioxide emissions, and oil prices: evidence from South America. *Energy Sources, Part B: Economics, Planning, and Policy*, *10*, 281–287. doi:https://doi.org/10.1080/15567249.2013.853713.

Apergis, N., Payne, J. E., Menyah, K., & Wolde-Rufael, Y. (2010). On the causal dynamics between emissions, nuclear energy, renewable energy, and economic growth. *Ecological Economics*, *69*, 2255–2260. doi:https://doi.org/10.1016/j.ecolecon.2010.06.014.

Apley, D. W., & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *82*, 1059–1086. doi:https://doi.org/10.1111/rssb.12377.

Bélaïd, F., & Youssef, M. (2017). Environmental degradation, renewable and non-renewable electricity consumption, and economic growth: Assessing the evidence from Algeria. *Energy policy*, *102*, 277–287. doi:https://doi.org/10.1016/j.enpol.2016.12.012.

Belkin, M., Hsu, D., Ma, S., & Mandal, S. (2019). Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences*, *116*, 15849–15854. doi:https://doi.org/10.1073/pnas.1903070116.

Beltrami, F., Fontini, F., & Grossi, L. (2021). The value of carbon emission reduction induced by renewable energy sources in the Italian power market. *Ecological Economics*, *189*, 107149. doi:https://doi.org/10.1016/j.ecolecon.2021.107149.

Bento, J. P. C., & Moutinho, V. (2016). CO2 emissions, non-renewable and renewable electricity production, economic growth, and international trade in Italy. *Renewable and Sustainable Energy Reviews*, *55*, 142–155. doi:https://doi.org/10.1016/j.rser.2015.10.151.

Bersani, A. M., Falbo, P., & Mastroeni, L. (2022). Is the ETS an effective environmental policy? Undesired interaction between energy-mix, fuel-switch and electricity prices. *Energy Economics*, (p. 105981). doi:https://doi.org/10.1016/j.eneco.2022.105981.

Bilgili, F., Koçak, E., & Bulut, Ü. (2016). The dynamic impact of renewable energy consumption on CO2 emissions: a revisited Environmental Kuznets Curve approach. *Renewable and Sustainable Energy Reviews*, *54*, 838–845. doi:https://doi.org/10.1016/j.rser.2015.10.080.

Boehmke, B., & Greenwell, B. (2019). *Hands-on machine learning with R*. Chapman and Hall/CRC. doi:https://doi.org/10.1201/9780367816377.

Boontome, P., Therdyothin, A., & Chontanawat, J. (2017). Investigating the causal relationship between non-renewable and renewable energy consumption, CO2 emissions and economic growth in Thailand. *Energy Procedia*, *138*, 925–930. doi:https://doi.org/10.1016/j.egypro.2017.10.141.

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5–32. doi:https://doi.org/10.1023/A:1010933404324.

Brouwer, A. S., van den Broek, M., Zappa, W., Turkenburg, W. C., & Faaij, A. (2016). Least-cost options for integrating intermittent renewables in low-carbon power systems. *Applied Energy*, *161*, 48–74. doi:https://doi.org/10.1016/j.apenergy.2015.09.090.

Bulut, U. (2017). The impacts of non-renewable and renewable energy on CO 2 emissions in Turkey. *Environmental Science and Pollution Research*, *24*, 15416–15426. doi:https://doi.org/10.1007/s11356-017-9175-2.

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* KDD '16 (pp. 785–794). New York, NY, USA: ACM. URL: http://doi.acm.org/10.1145/2939672.2939785. doi:https://doi.org/10.1145/2939672.2939785.

Chen, Y., Wang, Z., & Zhong, Z. (2019). CO2 emissions, economic growth, renewable and non-renewable energy production and foreign trade in China. *Renewable energy*, *131*, 208–216. doi:https://doi.org/10.1016/j.renene.2018.07.047.

Chu, C.-T., & Hawkes, A. D. (2020). Optimal mix of climate-related energy in global electricity systems. *Renewable Energy*, *160*, 955–963. doi:https://doi.org/10.1016/j.renene.2020.06.104.

Conover, W. J. (1999). *Practical nonparametric statistics* volume 350. john wiley & sons.

Dogan, E., & Seker, F. (2016). Determinants of CO2 emissions in the European Union: the role of renewable and non-renewable energy. *Renewable Energy*, *94*, 429–439. doi:https://doi.org/10.1016/j.renene.2016.03.078.

Dong, K., Sun, R., & Hochman, G. (2017). Do natural gas and renewable energy consumption lead to less CO2 emission? Empirical evidence from a panel of BRICS countries. *Energy*, *141*, 1466–1478. doi:https://doi.org/10.1016/j.energy.2017.11.092.

Débat national sur la transition énergétique (2013). Synthèse des travaux du débat national sur la transition énergétique de la France.

Edenhofer, O. (2015). *Climate change 2014: mitigation of climate change* volume 3. Cambridge University Press.

Engeland, K., Borga, M., Creutin, J.-D., François, B., Ramos, M.-H., & Vidal, J.-P. (2017). Space-time variability of climate variables and intermittent renewable electricity production–A review. *Renewable and Sustainable Energy Reviews*, *79*, 600–617. doi:https://doi.org/10.1016/j.rser.2017.05.046.

European Commission (2021). CO2 emissions from energy use clearly decreased in the EU in 2020. URL: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20210507-1 accessed on 12 February 2022.

Farhani, S., & Shahbaz, M. (2014). What role of renewable and non-renewable electricity consumption and output is needed to initially mitigate CO2 emissions in MENA region? *Renewable and Sustainable Energy Reviews*, *40*, 80–90. doi:https://doi.org/10.1016/j.rser.2014.07.170.

Fisher, A., Rudin, C., & Dominici, F. (2019). All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. *J. Mach. Learn. Res.*, *20*, 1–81. doi:https://www.jmlr.org/papers/volume20/18-760/18-760.pdf.

François, B., Hingray, B., Raynaud, D., Borga, M., & Creutin, J. (2016). Increasing climate-related-energy penetration by integrating run-of-the river hydropower to wind/solar mix. *Renewable Energy*, *87*, 686–696. doi:https://doi.org/10.1016/j.renene.2015.10.064.

Gernaat, D. E., de Boer, H. S., Daioglou, V., Yalew, S. G., Müller, C., & van Vuuren, D. P. (2021). Climate change impacts on renewable energy supply. *Nature Climate Change*, *11*, 119–125. doi:https://doi.org/10.1038/s41558-020-00949-9.

Goh, T., Ang, B., Su, B., & Wang, H. (2018). Drivers of stagnating global carbon intensity of electricity and the way forward. *Energy Policy*, *113*, 149–156. doi:https://doi.org/10.1016/j.enpol.2017.10.058.

Granger, C. W. J., & Newbold, P. (2014). *Forecasting economic time series*. Academic Press.

Hamilton, N. E., & Ferry, M. (2018). ggtern: Ternary diagrams using ggplot2. *Journal of Statistical Software*, *87*, 1–17. doi:https://doi.org/10.18637/jss.v087.c03.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D. et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*, 1999–2049. doi:https://doi.org/10.1002/qj.3803.

Ho, L. T., Dubus, L., De Felice, M., & Troccoli, A. (2020). Reconstruction of multidecadal country-aggregated hydro power generation in Europe based on a random forest model. *Energies*, *13*, 1786. doi:https://doi.org/10.3390/en13071786.

International Energy Agency (IEA) (2020). European Union 2020 Energy Policy Review. URL: https://www.iea.org/reports/european-union-2020 accessed on 13 February 2022.

Jaforullah, M., & King, A. (2015). Does the use of renewable energy sources mitigate CO2 emissions? A reassessment of the US evidence. *Energy Economics*, *49*, 711–717. doi:https://doi.org/10.1016/j.eneco.2015.04.006.

Jebli, M. B., & Youssef, S. B. (2015). The environmental Kuznets curve, economic growth, renewable and non-renewable energy, and trade in Tunisia. *Renewable and Sustainable Energy Reviews*, *47*, 173–185. doi:https://doi.org/10.1016/j.rser.2015.02.049.

Jebli, M. B., & Youssef, S. B. (2017). The role of renewable energy and agriculture in reducing CO2 emissions: Evidence for North Africa countries. *Ecological indicators*, *74*, 295–301. doi:https://doi.org/10.1016/j.ecolind.2016.11.032.

Jebli, M. B., Youssef, S. B., & Ozturk, I. (2016). Testing environmental Kuznets curve hypothesis: The role of renewable and non-renewable energy consumption and trade in OECD countries. *Ecological Indicators*, *60*, 824–831. doi:https://doi.org/10.1016/j.ecolind.2015.08.031.

Karmellos, M., Kopidou, D., & Diakoulaki, D. (2016). A decomposition analysis of the driving factors of CO2 (Carbon dioxide) emissions from the power sector in the European Union countries. *Energy*, *94*, 680–692. doi:https://doi.org/10.1016/j.energy.2015.10.145.

Khan, M. T. I., Ali, Q., & Ashfaq, M. (2018). The nexus between greenhouse gas emission, electricity production, renewable energy and agriculture in Pakistan. *Renewable Energy*, *118*, 437–451. doi:https://doi.org/10.1016/j.renene.2017.11.043.

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software, Articles*, *28*, 1–26. URL: https://www.jstatsoft.org/v028/i05. doi:https://doi.org/10.18637/jss.v028.i05.

Leamer, E. E. (1985). Vector autoregressions for causal inference? In *Carnegie-rochester conference series on Public Policy* (pp. 255–304). Elsevier volume 22. doi:https://doi.org/10.1016/0167-2231(85)90035-1.

Lehner, B., Czisch, G., & Vassolo, S. (2005). The impact of global change on the hydropower potential of Europe: a model-based analysis. *Energy Policy*, *33*, 839–855. doi:https://doi.org/10.1016/j.enpol.2003.10.018.

Leitão, N. C. (2014). Economic growth, carbon dioxide emissions, renewable energy and globalization. *International Journal of Energy Economics and Policy*, *4*, 391–399.

Liu, X., Zhang, S., & Bae, J. (2017a). The impact of renewable energy and agriculture on carbon dioxide emissions: investigating the environmental Kuznets curve in four selected ASEAN countries. *Journal of cleaner production*, *164*, 1239–1247. doi:https://doi.org/10.1016/j.jclepro.2017.07.086.

Liu, X., Zhang, S., & Bae, J. (2017b). The nexus of renewable energy-agriculture-environment in BRICS. *Applied energy*, *204*, 489–496. doi:https://doi.org/10.1016/j.apenergy.2017.07.077.

Long, X., Naminse, E. Y., Du, J., & Zhuang, J. (2015). Nonrenewable energy, renewable energy, carbon dioxide emissions and economic growth in China from 1952 to 2012. *Renewable and Sustainable Energy Reviews*, *52*, 680–688. doi:https://doi.org/10.1016/j.rser.2015.07.176.

Menyah, K., & Wolde-Rufael, Y. (2010). CO2 emissions, nuclear energy, renewable energy and economic growth in the US. *Energy policy*, *38*, 2911–2915. doi:https://doi.org/10.1016/j.enpol.2010.01.024.

Ministère de la Transition écologique (2019). l'Assemblée nationale inscrit la neutralité carbone et l' urgence écologique et la crise climatique dans la loi. URL: https://www.ecologie.gouv.fr/lassemblee-nationale-inscrit-neutralite-carbone-et-l-urgence-ecologique-et-crise-climatique-dans-loi (Accessed on 14 February 2022).

Molnar, C. (2020). *Interpretable machine learning*. Lulu. com. URL: https://christophm.github.io/interpretable-ml-book/.

Molnar, C., Casalicchio, G., & Bischl, B. (2018). iml: An R package for interpretable machine learning. *Journal of Open Source Software*, *3*, 786. doi:https://doi.org/10.21105/joss.00786.

Möst, D., & Fichtner, W. (2010). Renewable energy sources in European energy supply and interactions with emission trading. *Energy Policy*, *38*, 2898–2910. doi:https://doi.org/10.1016/j.enpol.2010.01.023.

Özbuğday, F. C., & Erbas, B. C. (2015). How effective are energy efficiency and renewable energy in curbing CO2 emissions in the long run? A heterogeneous panel data analysis. *Energy*, *82*, 734–745. doi:https://doi.org/10.1016/j.energy.2015.01.084.

Paramati, S. R., Sinha, A., & Dogan, E. (2017). The significance of renewable energy use for economic output and environmental protection: evidence from the Next 11 developing economies. *Environmental Science and Pollution Research*, *24*, 13546–13560. doi:https://doi.org/10.1007/s11356-017-8985-6.

Qi, T., Zhang, X., & Karplus, V. J. (2014). The energy and CO2 emissions impact of renewable energy development in China. *Energy Policy*, *68*, 60–69. doi:https://doi.org/10.1016/j.enpol.2013.12.035.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria. URL: https://www.R-project.org/.

del Río, P. (2017). Why does the combination of the European Union Emissions Trading Scheme and a renewable energy target makes economic sense? *Renewable and Sustainable Energy Reviews*, *74*, 824–834. doi:https://doi.org/10.1016/j.rser.2017.01.122.

Rodrigues, J. F., Wang, J., Behrens, P., & de Boer, P. (2020). Drivers of CO2 emissions from electricity generation in the European Union 2000–2015. *Renewable and Sustainable Energy Reviews*, *133*, 110104. doi:https://doi.org/10.1016/j.rser.2020.110104.

Rogelj, J., Shindell, D., Jiang, K., Fifita, S., Forster, P., Ginzburg, V., Handa, C., Kheshgi, H., Kobayashi, S., Kriegler, E. et al. (2018). Mitigation pathways compatible with 1.5 C in the context of sustainable development. In *Global warming of 1.5 C* (pp. 93–174). Intergovernmental Panel on Climate Change.

Réseau de Transport d'Électricité (RTE) (2020). Bilan des émissions de CO2. URL: https://bilan-electrique-2020.rte-france.com/production-emissions-de-co2/ accessed on 13 February 2022.

Sadorsky, P. (2009). Renewable energy consumption, CO2 emissions and oil prices in the G7 countries. *Energy Economics*, *31*, 456–462. doi:https://doi.org/10.1016/j.eneco.2008.12.010.

Saidi, K., & Mbarek, M. B. (2016). Nuclear energy, renewable energy, CO2 emissions, and economic growth for nine developed countries: Evidence from panel Granger causality tests. *Progress in Nuclear Energy*, *88*, 364–374. doi:https://doi.org/10.1016/j.pnucene.2016.01.018.

Saint-Drenan, Y.-M., Wald, L., Ranchin, T., Dubus, L., & Troccoli, A. (2018). An approach for the estimation of the aggregated photovoltaic power generated in several European countries from meteorological data. *Advances in Science and Research*, *15*, 51–62. doi:https://doi.org/10.5194/asr-15-51-2018.

Salazar-Núñez, H. F., Venegas-Martínez, F., & Lozano-Díez, J. A. (2021). Assessing the interdependence among renewable and non-renewable energies, economic growth, and CO2 emissions in Mexico. *Environment, Development and Sustainability*, (pp. 1–17). doi:https://doi.org/10.1007/s10668-021-01968-y.

Sebri, M., & Ben-Salha, O. (2014). On the causal dynamics between economic growth, renewable energy consumption, CO2 emissions and trade openness: Fresh evidence from BRICS countries. *Renewable and Sustainable Energy Reviews*, *39*, 14–23. doi:https://doi.org/10.1016/j.rser.2014.07.033.

Seyedhashemi, H., Hingray, B., Lavaysse, C., & Chamarande, T. (2021). The Impact of Low-Resource Periods on the Reliability of Wind Power Systems for Rural Electrification in Africa. *Energies*, *14*, 2978. doi:https://doi.org/10.3390/en14112978.

Shafiei, S., & Salim, R. A. (2014). Non-renewable and renewable energy consumption and CO2 emissions in OECD countries: a comparative analysis. *Energy Policy*, *66*, 547–556. doi:https://doi.org/10.1016/j.enpol.2013.10.064.

Sharif, A., Mishra, S., Sinha, A., Jiao, Z., Shahbaz, M., & Afshan, S. (2020). The renewable energy consumption-environmental degradation nexus in Top-10 polluted countries: Fresh insights from quantile-on-quantile regression approach. *Renewable Energy*, *150*, 670–690. doi:https://doi.org/10.1016/j.renene.2019.12.149.

Shirizadeh, B., Perrier, Q., & Quirion, P. (2022). How sensitive are optimal fully renewable power systems to technology cost uncertainty? *The Energy Journal*, *43*. doi:https://doi.org/10.5547/01956574.43.1.bshi.

Shirizadeh, B., & Quirion, P. (2021). Low-carbon options for the French power sector: What role for renewables, nuclear energy and carbon capture and storage? *Energy Economics*, *95*, 105004. doi:https://doi.org/10.1016/j.eneco.2020.105004.

Sinha, A., Shahbaz, M., & Sengupta, T. (2018). Renewable energy policies and contradictions in causality: a case of Next 11 countries. *Journal of cleaner production*, *197*, 73–84. doi:https://doi.org/10.1016/j.jclepro.2018.06.219.

Thakur, A. (2020). *Approaching (almost) any machine learning problem*. Abhishek Thakur.

Waheed, R., Chang, D., Sarwar, S., & Chen, W. (2018). Forest, agriculture, renewable energy, and CO2 emission. *Journal of Cleaner Production*, *172*, 4231–4238. doi:https://doi.org/10.1016/j.jclepro.2017.10.287.

Waisman, H., De Coninck, H., & Rogelj, J. (2019). Key technological enablers for ambitious climate goals: insights from the IPCC special report on global warming of 1.5 C. *Environmental Research Letters*, *14*, 111001. doi:https://doi.org/10.1088/1748-9326/ab4c0b.

Wilcoxon, F. (1992). Individual comparisons by ranking methods. In *Breakthroughs in statistics* (pp. 196–202). Springer. doi:https://doi.org/10.1007/978-1-4612-4380-9_16.

World Nuclear Association (2022). Nuclear Power in France. URL: https://www.world-nuclear.org/information-library/country-profiles/countries-a-f/france.aspx accessed on 13 February 2022.

Zeb, R., Salar, L., Awan, U., Zaman, K., & Shahbaz, M. (2014). Causal links between renewable energy, environmental degradation and economic growth in selected SAARC countries: progress towards green economy. *Renewable energy*, *71*, 123–132. doi:https://doi.org/10.1016/j.renene.2014.05.012.