Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Real-time CO2 emissions estimation in Spain and application to the COVID-19 pandemic

Luis F.S. Merchante a, b, Delia Clar a, Alberto Carnicero a, b, Francisco J. Lopez-Valdes a, b, *, Jesús R. Jimenez-Octavio a, b

a ICAI School of Engineering, 25 Alberto Aguilera, 28015, Madrid, Spain
b IIT Instituto de Investigacion Tecnologica, Universidad Pontificia Comillas, Sta Cruz de Marcenado 26, 28015, Madrid, Spain

A R T I C L E   I N F O

Article history:
Received 29 October 2020
Received in revised form 3 February 2021
Accepted 15 February 2021
Available online 20 February 2021

MSC: 00e02 62e07

Keywords: Forecast COVID-19 CO2 Emissions

A B S T R A C T

CO2 emissions are one of the major contributors to global warming. The variety of emission sources and the nature of CO2 hinders estimating its concentration in real time and therefore to adopt flexible policies that contribute to its control and, ultimately, to reduce its effects. Spain is not exempted from this challenge and CO2 emissions are published only at the end of the year and as an aggregated value for the whole country, without recognising the existing differences between the regions (the so-called, Autonomous Communities). The recent COVID-19 pandemic is a clear example of the need of accurate and fast estimation methods so that policies can be tailored to the current status and not to a past one. This paper provides a method to estimate monthly emissions of CO2 for each AACC in Spain based on data that are published monthly by the relevant administrations. The paper discusses the approximations needed in the development of the method, predicts the drop in emissions due to the reduced industrial activity during the pandemic in Spain and provides the estimation of future emissions under three recovery scenarios after the pandemic.

© 2021 Elsevier Ltd. All rights reserved.

1. Introduction

Measuring greenhouse gas emissions has been one of the main concerns of many governments for the last decades. Even if methane has the most potential in contribution to global warming, carbon dioxide (CO2) currently ranks first in affecting global warming due to its abundance in the atmosphere. In addition, CO2 is the primary greenhouse gas (GHG) originated from human activities (United States Environmental Protection Agency, 2019). The Environmental Protection Agency of the United States and the most recent report from the Mitigation of Climate Change working group of the Intergovernmental Panel on Climate Change (IPCC) of the United Nations identified that the vast majority of anthropogenic carbon dioxide emissions come from combustion of fossil fuels (principally coal, oil, and natural gas), with additional contributions coming from deforestation, changes in land use, soil erosion and agriculture (including livestock) (United States Environmental Protection Agency, 2019; Mitigation of Climate Change working group of the Intergovernmental Panel on Climate Change, 2018).

Since 1970, CO2 emissions have increased by about 90%, with emissions from fossil fuel combustion and industrial processes contributing to about 78% of the total GHG emissions increase from 1970 to 2011. The variety of sources and the nature of the non-anthropogenic CO2 that is naturally present on the atmosphere difficult estimating its concentration in real time (Le Quéré et al., 2020).

Consequently, CO2 emission values are normally released in an aggregated manner at the end of each year. When instantaneous concentration of CO2 emission are required, estimations must be made using proxy data which could be available almost at real time. These estimations are usually based on satellite images (Doll et al., 2000; Meng et al., 2014; Ghosh et al.; Shi et al., 2016; Nassar et al., 2017) but other approaches use proxy variables such as the fractional change in activity levels for each sector (Le Quéré et al., 2020).
or other socio-economic variables (Begum et al., 2015; Hong et al., 2018) to estimate the instantaneous concentration of CO2 emissions. In this regard, a great variety of short-term and long-term forecasting techniques have been used to estimate GHG emissions. A useful review can be found in Table 1 from reference Hong et al. (2018). Most of those techniques are based on Evolutionary Algorithms (Karabulut et al., 2008; Mousavi et al., 2014; Fang et al., 2018), although Artificial Networks are also very popular on this domain (Behrang et al., 2011; Kankal et al., 2011; Ardakani and Ardeshali, 2014; Guo et al., 2018; Heydari et al., 2019). Fewer references can be found on Support Vector Machines (Sun and Liu, 2016; Saleh et al., 2016; Ahmadi et al., 2019) or Regressions (Köne and Bükö, 2010; Azadeh et al., 2017; Hosseini et al., 2019). Not many studies have been found testing other Machine Learning techniques, especially those based on ensemble methods (Dietterich, 2000) like Random Forest (Wei et al., 2018), Adaboost (Zhou et al., 2016; del Rio-Chanona et al., 2020). These restrictions met the cant impact in other areas, like in the environmental objective of reducing the virus spread, but they also had a signi

Table 1

| Community                  | Description                                                                 | Main features                                         |
|----------------------------|-----------------------------------------------------------------------------|-------------------------------------------------------|
| Linear Regressor           | Fits a linear model to minimize the quadratic mean squared error             | States a simple base line of accuracy                  |
| K-Nearest Neighbors Regressor | The value is predicted by local interpolation of the nearest data in the k-neighbourhood | Simple, non-parametric, robust to noisy data          |
| Decision Trees Regressor   | Non-parametric method by learning decision rules resulting in local linear regressions | Non-parametric, interpretable                          |
| Random Forest Regressor    | Ensemble of Decision Trees to improve generalizability                       | More resistant to overfitting, very stable             |
| Gradient Boosting Regressor| Ensemble of weak decision trees models. New predictors are fitted with mistakes | Reduce bias and variance                               |
| Epsilon Support Vector Regressor | Fit a hyper-plane to the data transformed by a RBF kernel. Some error controlled by epsilon is tolerated | Computational complexity does not depend on the input data dimensionality |
| Kernel Ridge Regressor     | Combines ordinary Least Squares with L2 penalty on the coefficients with kernel trick | Efficient non-linear fitting                           |

2. Data and methodology

2.1. Data

Publicly available data per each Spanish AACC in the period 2011–2018 were retrieved from several sources (see Table 2). Feature selection techniques were applied, so that the CO2 emissions model for each AACC could use different predictors, in order to contribute to reveal the nature of CO2 emissions within each one. The list of variables that were explored to model the CO2 emissions in the different Autonomous Communities are listed in Table 2.

2.2. Apportioning yearly reported variables into monthly figures

The most recent information about CO2 emissions was retrieved from reports from the Ministry for Ecological Transition and Demographic Challenge (Ministerio para la Transición Ecológica y el Reto Demográfico, 2020), where equivalent emissions for AACC are provided from 1990 to 2018. It should be noted that the volume of CO2 is provided only annually. Since the model to be developed in this study seeks to predict the emissions on a monthly basis (to be able to detect the impact of rapid events such as the COVID-19 crisis), this variable needed to be transformed into monthly figures before being able to include it in the model.

To this end, an approach based on monthly reported energy indicators was followed. These indicators are published by Red Elétrica Española (REE, the only high voltage electric transport operator in Spain) (Red Eléctrica Española, 2020), REE reports on the monthly emissions produced by non-renewable energy generation in Spain and on the energy produced monthly by every AACC, desegregated by energy type. The approximation is that the monthly distribution found on the emissions from non-renewable energy generation applies also to the global anthropogenic CO2 monthly distribution. This hypothesis is supported by a very similar

1 Autonomous Communities is a level of political and administrative division similar, to a greater or lesser extent, to the French Departments or the Landers in Germany.
behaviour between both annual time series as depicted in Fig. 1. If their annual values are closely correlated, it has sense finding the same behaviour on monthly distributions.

The approach followed to apportion in monthly figures the yearly amount of CO2 emissions is detailed in Algorithm 1.

**Algorithm 1.** Approach to apportion yearly reported emissions of CO2 in monthly figures

1. Download global annually CO2 anthropogenic emissions by AACC
2. Download values of energy structure generation for non-renewable sources per month and AACC and compute their monthly contribution to the national figures as a percentage
3. Download monthly CO2 emissions for non-renewable energy generation only available with national and peninsular aggregation, islands and autonomous cities (Ceuta and Melilla)
4. Apply the calculated monthly percentages of energy generation contribution for every AACC to the national aggregation of CO2 emissions for non-renewable energy generation
5. Compute the percentage of contribution of monthly CO2 emissions for non-renewable energy generation per AACC to the national aggregated actual values
6. Apply those percentages to the actual annual anthropogenic CO2 values available at AACC level to obtain the monthly approximation
2.3. Model construction, training and validation

In this study, it was considered that regressions techniques and, more precisely, regression meta-algorithms based on ensembles could provide useful models without sacrificing accuracy. Several of those regression techniques were tested and compared by means of the $R^2$ coefficient, thanks to recent implementations of Machine Learning libraries that allowed the authors to train numerous modeling techniques with few efforts (Pedregosa et al., 2011; Pytorch, 2020; TensorFlow, 2020). The interpretation of the $R^2$ coefficient is that the closer the $R^2$ value is to one, the more accurate the prediction is. In this context, negative $R^2$ values mean that the average of the data provides a better fit to the outcomes than the predicted values.

Before training, data were standardized (removing the mean and scaling to unit variance), as many regression techniques assume a Gaussian distribution of the attributes. This process is not required for regression algorithms that do not make this Gaussian assumption (e.g. Decision Trees), but it is mandatory for several others. The regression techniques used in this study were: K-Nearest Neighbors Regressor, Decision Trees, Random Forest and Gradient Boosting Regressors, Epsilon-Support Vector Regression, Linear and Kernel Ridge Regressors. Table 1 shows the main features for each technique. Modelling was combined with feature selection (by means of Sequential Forward Floating Selection (Pudil et al., 1994)) and hyper-parameter searching techniques to find the best fit. All the models were trained with data from years 2011–2016, and validated with data from 2017 to 2018. The process described above is resumed in Algorithm 2. It is only after training and validation that the model can be used to predict the emissions in 2019 and 2020.

**Algorithm 2. Method followed for modelling CO2 emissions**

1. Build train matrix with data from 2011 to 2018. This train matrix is limited to 2018 due to the availability of validated CO2 emissions anthropogenic emissions
2. Split train matrix in train (data from 2011 to 2016) and test (2017 to 2018)
3. The modelling process involves a pipeline with standardization, feature selection, model parametrization, cross validation and score calculation
4. Previous process is tested for different regression techniques and parameter grids
5. Best models and parameters are selected
6. Models are validated against test data
7. Models can be used to forecast 2019 and 2020

2.4. Scenarios used to predict future CO2 emissions

Finally, the model obtained was used to predict the change in emission levels in three hypothetical scenarios that can potentially occur associated to the recovery after the COVID-19 pandemic.

At the time of writing this document, the Statistic National Institute of Spain had estimated a –21.5% Gross Domestic Product (GDP) fall for the 2nd quarter of 2020 (flash estimate) (Instituto Nacional de Estadística, 2020). Based on this prediction, this study hypothesized that, due to the pandemic, the GDP in Spain remained at this level until the end of 2020. Under this assumption, three potential recovery scenarios were simulated, as follows:

- Scenario 1, V-shape recovery: economic activity will get back to the level of January 2020 in a linear fashion by January 2022.
- Scenario 2, Slow V-shape recovery: similar to the previous one, but full recovery will be reached by January 2023.
- Scenario 3, U-shape recovery: in which the economic activity in January 2022 would be similar to that of January 2021, rising back to the values of January 2020 by January 2023.
Relevant variables per model.

Table 4
Average and standard deviation values of the $R^2$ coefficient for the different models tested.

| Community      | linear avg | linear std | knn avg | knn std | decision Trees avg | decision Trees std | random Forest avg | random Forest std | gradient Boosting avg | gradient Boosting std | svr avg | svr std | krr avg | krr std |
|----------------|------------|------------|---------|---------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|---------|---------|---------|---------|
| Andalucia      | -1.04      | 1.46       | 0.17    | 0.19    | 0.12              | 0.34              | 0.60              | 0.13              | 0.51              | 0.20              | -0.20   | 0.23    | -26.24 | 10.22   |
| Aragon         | -31.26     | 60.04      | -0.33   | 0.79    | -0.63             | 1.10              | 0.01              | 0.45              | 0.02              | 0.39              | -0.25   | 0.25    | -15.71 | 14.45   |
| Cantabria      | -1.63      | 1.77       | -0.28   | 0.38    | -1.00             | 1.12              | -0.18             | 0.22              | -0.37             | 0.39              | -0.13   | 0.12    | -26.94 | 12.50   |
| Castilla la Mancha | -2.87   | 4.92       | -0.04   | 0.32    | -1.17             | 1.19              | -0.02             | 0.17              | -0.04             | 0.28              | -0.24   | 0.28    | -21.16 | 14.05   |
| Castilla y Leon | -2.23      | 3.18       | 0.07    | 0.44    | -1.84             | 2.08              | -0.28             | 0.71              | -0.63             | 1.02              | -0.17   | 0.20    | -7.04  | 9.21    |
| Cataluña       | -1.85      | 3.15       | 0.14    | 0.30    | -0.44             | 0.82              | 0.14              | 0.31              | 0.11              | 0.35              | -0.20   | 0.17    | -5.46  | 21.47   |
| País Vasco     | -1.03      | 1.82       | -0.00   | 0.38    | -0.46             | 0.69              | 0.26              | 0.17              | 0.11              | 0.10              | -0.12   | 0.10    | -16.25 | 6.54    |
| Principado de Asturias | -2.14   | 3.28       | 0.18    | 0.31    | -0.72             | 1.10              | 0.24              | 0.28              | 0.21              | 0.49              | -0.11   | 0.12    | -10.31 | 5.72    |
| Comunidad de Madrid | -6.86   | 6.12       | -0.07   | 0.41    | -0.40             | 0.43              | 0.10              | 0.21              | 0.15              | 0.31              | -0.14   | 0.16    | -90.52 | 44.94   |
| Comunidad de Navarra | -3.09   | 5.39       | 0.90    | 5.16    | 0.06             | 0.85              | -0.13             | 0.75              | -0.08             | 0.09              | -18.38  | 12.18   |
| Comunidad Valenciana | -4.17   | 7.32       | 0.00    | 0.47    | -0.62             | 0.66              | 0.02              | 0.27              | 0.04              | 0.39              | -0.07   | 0.09    | -29.39 | 18.45   |
| Extremadura    | -0.32      | 1.32       | -0.22   | 0.83    | -1.49             | 1.31              | -0.42             | 0.53              | -0.77             | 0.71              | -0.10   | 0.12    | -15.91 | 8.55    |
| Galicia        | -2.59      | 5.34       | 0.24    | 0.24    | -0.03             | 0.24              | 0.42              | 0.12              | 0.39              | 0.14              | -0.13   | 0.16    | -10.49 | 2.18    |
| Islas Baleares | -3.03      | 5.38       | 0.45    | 0.47    | 0.28              | 0.38              | 0.58              | 0.25              | 0.49              | 0.31              | -0.31   | 0.40    | -36.95 | 4.01    |
| Islas Canarias | -0.70      | 13.21      | 0.15    | 0.28    | -0.49             | 0.51              | 0.25              | 0.22              | 0.08              | 0.36              | -0.14   | 0.25    | -66.52 | 16.03   |
| La Rioja       | -11.90     | 15.68      | -0.04   | 0.29    | -0.77             | 0.40              | 0.05              | 0.32              | -0.14             | 0.47              | -0.08   | 0.10    | -1.39  | 0.48    |
| Región de Murcia | -3.79    | 6.53       | -0.18   | 0.47    | -0.33             | 0.63              | 0.19              | 0.25              | 0.18              | 0.23              | -0.14   | 0.14    | -21.17 | 13.26   |

1. gdp p.c. stands for gdp per capita.
2. $tCO_2$eq stands for CO2 emissions from non-renewable electric generation.
3. trx_inmob stands for number of real state operations.
4. Services idx stands for index of activity of service sector.

Table 5
R-squared values per model (from best to worse accuracy).

| Community    | Accuracy ($R^2$) |
|--------------|-----------------|
| Galicia      | 0.929011        |
| Castilla la Mancha | 0.902679 |
| Cataluña    | 0.858921        |
| Andalucia   | 0.798672        |
| Principado de Asturias | 0.796600 |
| Aragón      | 0.772803        |
| Castilla y Leon | 0.758328 |
| Islas Baleares | 0.749662 |
| Islas Canarias | 0.735272 |
| Comunidad de Madrid | 0.684010 |
| Comunidad de Navarra | 0.631332 |
| La Rioja    | 0.559210        |
| Extremadura | 0.459273        |
| Comunidad Valenciana | 0.394295 |
| País Vasco  | 0.211449        |
| Región de Murcia | -0.218722 |
| Cantabria   | -2.792710       |

3. Results
3.1. Model selection and results validation

As aforementioned, all the algorithms from Table 1 were trained and tested using 50 repetitions of a 5-fold cross-validation process. Each cross-validated training returned five $R^2$ whose mean and standard deviation were averaged over the 50 repetitions. Table 3 shows the average and standard deviation values of the $R^2$ coefficient. Best values were achieved by Random Forest and Gradient Boosting Regressors. Eventually a Gradient Boosting Regressor (Friedman, 1999, 2000) was chosen because its implementation allowed us to obtain a measure of uncertainty using lower and upper prediction intervals (Scikit-learn, 2020). As mentioned before, this technique combined with feature selection allowed the authors to identify the most relevant factors that explained CO2 emissions within each AACC. These variables are shown in Table 4.

As shown in Table 4, the models are from different complexity as intended with the proposed feature selection methodology. The
Fig. 2. Actual data (kTmCO₂eq) along with predicted data for 2017 and 2018 and forecast values for 2019 and 2020 for every AACC models.
Results also show that emissions from non-renewable electric generation are possibly the best predictor of CO₂ emissions because the former variable (tCO₂eq) was selected by all models. The accuracy of the model can be quantified by means of the $R^2$ coefficient as shown in Table 5.

3.2. Predicted CO₂ emissions in each AACC in Spain in 2019 and 2020

Eventually, given the accuracy obtained in the validation, the model can be used to forecast CO₂ values for 2019 and 2020. These estimations are shown in Figs. 2 and 3. These figures show validation data from 2017 to 2018 and forecast data for 2019 and 2020. The red lines are the values predicted by the models and the shaded regions are bounded by the lower and upper limits representing the 10th and 90th percentiles.

CO₂ emission values estimated for 2019 and 2020 cannot be validated since no official figures have been published to date. However, the model predicts the decrease of emissions during 2019 since 58.6% of the electricity generated in Spain in 2019 did not emit CO₂ because it came from renewable sources. The prediction also shows a decrease in emissions in the first semester of 2020 that can possibly be linked to the COVID-19 pandemic.

3.3. Required pre-analysis for scenarios forecasting

The good fit of the model is strongly associated to the predictor quantifying the emissions from non-renewable electric generation.

Thus, the estimation of the anthropogenic CO₂ emitted in future scenarios requires computing the predictors based on the assumptions taken for each scenario, including estimating the amounts of non-renewable electric generation in these scenarios. To overcome this difficulty, a new set of models to predict the amount of CO₂ emitted from non-renewable energies per every AACC was developed using the same techniques as the ones described in the Data and methodology section.

To validate the accuracy of these new models per AACC, the model used to predict the emissions of CO₂ was run again using the prediction of CO₂ in the generation of non-renewable energy per AACC instead of the available reported data. As it could be expected, less precise predictions were obtained than with the first set of models. Table 6 shows the variables that entered in each model. More variables needed to be considered to get to a reasonable level of results. Figs. 4 and 5 show the estimation obtained.

Even if this new set of models are not extremely accurate, it was considered that they can provide a reasonable approximation to the monthly trends of CO₂ emissions for every AACC.

3.4. Scenarios

After the validation explained in the previous subsection, the trained model was used to predict the amount of CO₂ (kTmCO₂eq) that will be emitted to the atmosphere for each of the three scenarios described above until January 2023. Results are displayed in Fig. 6. Values from January 2019 until the moment of writing this document are colored in grey (estimations produced with actual

---

Table 6

| AACC                  | var1   | var2   | var3   | var4     | var5     | var6     | var7     | var8   | var9   | var10  |
|-----------------------|--------|--------|--------|----------|----------|----------|----------|--------|--------|--------|
| Andalucia             | month  | trx_immob | ss    | affl    | petrol_mov | ss     | affl     | petrol_indstr | deposits |
| Aragon                | month  | gbp    | gbp    | population | ss     | affl     | trax_immob | energy_demand | petrol_p.c   |
| Cantabria             | gbp    | gbp    | population | ss     | affl     | petrol_indstr | credits |
| Castilla la Mancha    | month  | prec  | gbp    | credits | ss     | affl     | trax_immob | energy_demand | petrol_p.c   |
| Cataluña              | month  | gbp    | gbp    | population | ss     | affl     | petrol_indstr | credits |
| Comunidad de Madrid    | month  | gbp    | gbp    | population | ss     | affl     | petrol_indstr | credits |
| Comunidad de Navarra  | month  | gbp    | gbp    | population | ss     | affl     | petrol_indstr | credits |
| Comunidad Valenciana  | month  | gbp    | gbp    | population | ss     | affl     | petrol_indstr | credits |
| Extremadura           | month  | gbp    | gbp    | population | ss     | affl     | petrol_indstr | credits |
| Galicia               | month  | gbp    | gbp    | population | ss     | affl     | petrol_indstr | credits |
| Islas Baleares        | gbp    | gbp    | energy_demand | ss     | affl     | petrol_indstr | credits |
| Islas Canarias        | gbp    | gbp    | population | ss     | affl     | petrol_indstr | credits |
| La Rioja              | gbp    | population | gbp    | population | ss     | affl     | petrol_indstr | credits |
| Region de Murcia      | gbp    | gbp    | population | ss     | affl     | petrol_indstr | credits |
|                       |        |        |        |          |          |          |          |        |        |        |

Fig. 3. Aggregated actual data (kTmCO₂eq) and predicted data for Spain.
Fig. 4. Actual data (TmCO$_2$eq) on electric non-renewable generation along with predicted data for 2017 and 2018 and forecasted values for 2019 and 2020 for every AACC models.
values of the predictors). Then, and until January 2023, the values have also been produced by the above method using hypothetical values of the predictors according to the different scenarios. Scenarios 1, 2 and 3 are colored in blue, orange or green respectively.

The scenarios described in section Scenarios used to predict future CO₂ emissions differed in the hypothesized recovery timeline of the GDP. The starting point is the second quarter of 2020 were the GDP was estimated to have fallen by 21.5% with respect the same quarter of 2019. Scenario 1 is the most optimistic of the three with a V-shape recovery; this scenario would reach by January 2022 the same levels of economic activity than those from January 2020. The remaining scenarios 2 and 3 predict V-shaped and U-shaped respectively slower recoveries, reaching the same economic activity values observed in January 2020 three years later, that is, in January 2023.

From the emissions point of view, the best scenario would be scenario 3 (green curve, slow U-shaped recovery) since the integral of its curve accumulates the least amount of kTmCO₂eq. Unfortunately, this scenario is also the worst from an economical perspective, as it assumes 12 months of recession at its lowest values and it is not until 2022 that the economy begins its recovery. On the contrary, the desired scenario for the economical recovery of Spain would be scenario 1 (blue curve) that is also the one with the largest amount of CO₂ emitted.

4. Discussion

The COVID-19 crisis has caused a decrease in pollutant emissions in many places in the world. It would be expected that once the crisis are over, emissions will return to their original levels. However, it can be argued that this situation can also lead to substantial shifts in energy efficiency and to the development of alternative, cleaner, energy sources (Le Quéré et al., 2020; Peters et al., 2011). In the case of COVID-19, for instance, it has been observed that the precautions taken to avoid infection had caused a decrease in the use of public transportation associated to an increase in the use of new means of clean personal mobility such as e-scooters or e-bikes. It should be noted that restrictions to contain the propagation of COVID-19 change very rapidly depending on the growth of the infection, even on a weekly or monthly basis. That means that the human and industrial activity may change dramatically in a very short period of time, making more difficult to estimate the CO₂ emissions using traditional methods. This is why the approximation shown in this work, even if it is only an approximation, can assist policy makers in predicting the impact in emissions of these restrictive policies.

Since AACC in Spain can implement policies to control CO₂ emissions in their territory, this paper used data retrieved at the AACC level so that the models developed here could be used by the regional authorities to design custom emission regulation policies that could be optimized for their individual characteristics. This approach resulted in training 19 different prediction models. From a modeling perspective, it is not a big issue. But having different models for each AACC complicated drawing conclusions about their similarities and differences regarding CO₂ emissions. Focusing on the variables that entered in the anthropogenic CO₂ and non-renewable CO₂ estimation models as shown in Tables 4 and 6, it was expected that the inclusion of the levels of CO₂ coming from non-renewable generation would be the key predictor for the 19 anthropogenic models. But the relevance of GDP related variables was also significant and somewhat less predictable. It seems that for those AACC with a quite predictable emission model of non-renewable CO₂, GDP and time of the year variables are enough to get an accurate prediction. For those communities with less predictable emissions, several extra variables related to financial products, state agent transactions or service sector levels were
required in addition. It should also be noted that the estimation of non-renewable generation emissions was more challenging, as shown by the increased number of variables required per model. Variables related to energy demand, the number of mortgages or affiliations to the social security become more relevant then. The information coded on those two tables could be of great value when designing emission reduction plans per AACC.

Two major difficulties were found: apportioning in monthly figures the yearly amount of CO2 emissions and the lack of mobility data available in Spain. To cope with the first one, a model to estimate these monthly figures was also developed. Unfortunately, despite the evident influence of traffic related emissions in the global CO2 emissions, the availability of information about mobility patterns per month and per AACC in Spain was very limited and did not allow us to establish any sensible method of estimating them. Initially, data related to the number of yearly crashes and victims were used as proxy variables for mobility, as there is literature pointing to the existing relationship between exposure and injuries (Segui-Gomez et al., 2011). However even if this data is available up to 2018 it did not end up contributing to the significance of the models. Some local administrations report the actual figures of private and public transportation use, unfortunately the number of informed cities is still too low to include this predictor in the models.

As mentioned above, one of the most significant predictors used in the model of anthropogenic CO2 emissions was the CO2 emissions produced in the generation of non-renewable electric energy. Recent publications describe energy generation as the second source of CO2 emissions in Spain (Ministry of Environment, 2020). However, generation values per AACC are not available and therefore published values could not be directly used as variables in the prediction models. To overcome this difficulty, we followed the approach described in the Data and methodology section that required the assumption of two hypotheses. The first one considers that the same distribution of energy per AACC applies to the CO2 emissions from the generation process. The second one considers that the monthly distribution of non-renewable emissions per AACC also applies to the monthly distribution of anthropogenic emissions. The first hypothesis seems quite reasonable. The second one is only supported on the similarity of the annual time-series that are the only actual data available. Even if this approach might present certain limitations, in the absence of other more detailed sources of data, it is the only way of estimating these values.

Although it is not possible to validate these assumptions, the model predicts the decrease of emissions due to new policies on contaminant energy production during 2019 and also due COVID-19 restrictions during 2020. These results were therefore consistent with reality. The strength of the approach is that, contrary to the total level of emissions that are only reported in a yearly bases, the exact amount of CO2 produced in generation is reported monthly by the relevant authority in Spain, and therefore can be used to predict the total amount of CO2 monthly emissions, which was the goal of this paper.

Last, one of the main contributions of this study is that the developed method allows quantifying economical recovery and anthropogenic emissions variables at the same time. The manuscript includes three hypothesized scenarios to show the feasibility of the method, but the method can be applied to other scenarios with different recovery lengths or more complex hypothesis affecting predictors to find an optimal solution between economic growth and keeping CO2 emissions at an affordable level, an issue that has been already identified in previous literature (Linares and Romero, 2000; Guerra et al., 2016; Lopez-Pena et al., 2012). It should be noted that the three studied scenarios were designed based on the estimation of the GDP reduction in Spain provided by the Statistic National Institute of Spain (Instituto Nacional de Estadística, 2020). The scenarios could be updated for other values of GDP change and the method developed here would serve as a tool to forecast interactions between economic recovery and CO2 emissions. For the proposed GDP reduction value, scenario 2 (orange curve in Fig. 6) might be a good compromise between reasonable recovery of economic activity and controlled emissions of CO2. However, when the economic survival of an entire country is at stake, economic recovery is likely a priority.

### 5. Conclusion

This study has proposed a method that estimates almost real time (monthly) CO2 emissions based on proxy variables related to energy production and consumption for each AACC in Spain. The method is flexible enough so that each AACC can forecast the short-term effect of implementing diverse energy policies in the CO2 emission levels. The method has been shown capable of capturing the effects of changes in regulation and of sudden events such as the COVID-19 pandemic. The model can also be used to produce a set of economic recovery scenarios after the pandemic, in which the participating variables can be estimated, to benchmark the effects of different energy policies in the emissions of CO2. The advantage of a data-centric policy definition is that, if the hypotheses about the predictors are revealed to be wrong, the scenarios might be updated with actual values whenever possible and policies can be dynamically improved.

Future work should include improving the complexity of the modeling techniques introducing theory of Recurrent Neural Networks that might capture better the correlation of the predictors and their evolution in time. For instance, the possibility of including mobility data at the AACC level and for a number of years can contribute largely to improving the predicting capabilities of the model.

### CRediT authorship contribution statement

- **Luis F.S. Merchante:** Conceptualization, Funding acquisition, Data curation, Supervision, Formal analysis.
- **Delia Clar:** Data curation, Alberto Carnicero: Conceptualization, Supervision, Formal analysis.
- **Francisco J. Lopez-Valdez:** Funding acquisition, Supervision, Formal analysis.
- **Jesús R. Jimenez-Octavio:** Conceptualization, Funding acquisition, Supervision, Formal analysis.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgements

This research is part of the project MoviliAD urbana post-covid19 (PP2020_06), funded by Universidad Pontificia Comillas. The authors are thankful for the support to carry out this research. The content of this manuscript is the solely responsibility of the authors and does not necessarily reflect the position of the funding institution.

### References

Ahmadi, M.H., Dehghani Madvar, M., Sadeghzadeh, M., Rezaei, M.H., Herrera, M., Shamshirband, S., 2019. Current status investigation and predicting carbon dioxide emission in Latin American countries by connectionist models. Energies...
Linares, P., Ardehali, M., 2014. Long-term electrical energy consumption forecasting for developing and developed economies based on different optimized models and historical data types. Energy 65, 452–461.

Azadeh, A., Jafari-Marandi, R., Abdollahi, M., Roudi, E., 2017. A novel benchmark methodology for estimating industrial electricity demand considering unsteady socio-economic conditions. Int. J. Bus. Perform. Manag. 13, 196–215.

Bartik, A.W., Bertrand, M., Cullen, Z., Claeser, E.L., Bu, L., Stanton, C., 2020. The impact of covid-19 on small business outcomes and expectations, 117, 17656–17660. https://doi.org/10.1016/j.prscn.2020.09.063.

Lofalipour, M.R., Falahi, M.A., Bastam, M., 2013. Prediction of co2 emissions in Iran using grey and arima models. Int. J. Energy Econ. Pol. 3, 229.

Meng, L., Graus, W., Worrell, E., Huang, B., 2014. Estimating co2 (carbon dioxide) emissions at urban scale using remote sensing tools (defense meteorological satellite program’s operational linescan system) nighttime light imagery: methodological challenges and a case study for China, 71, 468–478. https://doi.org/10.1016/j.enpol.2014.04.013.

Ministry for the Transition Ecologica and the Reto Demografico, 2020. Emisioes de gei por comunidades autónoma a partir do inventario espanol - serie 1990-2018. https://www.miteco.gob.es/calidad-y-evaluacion-ambiental/temas/sistema-espanol-de-inventario-emisionesgeoparquecerie1990-2018,

Ministry of Environment, 2020. Plan Nacional integrado de Energía y Clima (PNIEC) 2021-2030. Technical Report.

Mitigation of Climate Change working group of the Intergovernmental Panel on Climate Change, 2018. Drivers, Trends and Mitigation. https://www.ipcc.ch/site/assets/uploads/2018/02/IPCW3_AR5_chapter5.pdf. Accessed: 2020-08-30.

Mousavi, S.M., Mostalavi, E.S., Hosseinpur, F., 2014. Gene expression programming as a basis for new generation of electricity demand prediction models. Comput. Ind. Eng. 74, 120–128.

Nassar, R., Hill, T.G., McLinden, C.A., Wunch, D., Jones, D.B.A., Crisp, D., 2017. Quantifying co2 Emissions from Individual Power Plants from Space, vol. 44, 1–7. https://doi.org/10.1016/j.apenergy.2017.03.064.

Ozil, P. K. & Arun, T. ( ). Spillover of Covid-19: Impact on the Global Economy. doi: 10.2139/ssrn.3755270.

Pao, H.-T., Tsai, C.-M., 2011. Modeling and forecasting the co2 emissions, energy consumption, and economic growth in Taiwan, 36, 2019. doi: 10.1016/j.enpol.2011.01.032.

Pao, H.-T., Fu, H.-C., Tseng, C.-L., 2012. Forecasting of co2 emissions, energy consumption and economic growth in China using an improved grey model, 40, 400–409. doi: 10.1016/j.enpol.2011.02.031.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Peters, G.P., Minx, J.C., Weber, C.L., Edenhofe, O., 2011. Growth in emission transfers via international trade from 1990 to 2008. Proc. Natl. Acad. Sci. Unit. States Am. 108, 8903–8908.

Pudl, P., Novovicová, J., Kittler, J., 1994. Floating searches in feature selection. 5, 1119–1125. https://doi.org/10.1016/0166-218x(94)90042-7.

Pytorc, 2020. From Research to Production. https://pytorc.org/. Accessed: 2020-08-30.

Red Electric Espanola, 2020. Redata Api from Red Ellectrica. https://www.red.es/ajedrez/redes. Accessed: 2020-08-30.

Saleh, C., Dzikaiyullah, N.R., Nugroho, J.B., 2016. Carbon dioxide emission prediction using support vector machine. IOP Conf. Ser. Mater. Sci. Eng. 114, 012148. doi: 10.1088/1757-899X/114/1/012148.

Segui-Gomez, M., Lopez-Valdes, F.J., Guilleon-Grima, F., Smyth, E., Llorca, J., de Irala, J., 2012. The impact of socio-economic conditions on exposure to traffic from 2001 to 2010: A cross-sectional study. Environ. Sci. Pollut. Control Ser. 19, 28985–28997. https://doi.org/10.1016/j.envsci.2012.06.013.

Shirmohammadi, R., Aslani, A., Ghasempour, R., 2020. Challenges of carbon capture technologies development in developing countries. Sustainable Energy Technologies and Assessments 42, 100837.

Sun, W., Liu, M., 2016. Prediction and analysis of the three major industries and residential consumption co2 emissions based on least squares support vector machine in China, J. Clean. Prod. 122, 144–153.

TensorFlow, 2020. Tensorflow, the Deep Learning Library for python. https://www.tensorflow.org/. Accessed: 2020-08-30.

United States Environmental Protection Agency, 2019. Global Greenhouse Gas Emissions Data. https://www.epa.gov/energy/greenhouse-gas-emissions-data. Accessed: 2020-08-30.

Wei, S., Yuwei, W., Chongzhong, Z., 2018. Forecasting co2 emissions in hebei, China, through moth-flame optimization based on the random forest and extreme learning machine. Environ. Sci. Polit. Control Ser. 25, 28985–28997.

Worldmeter, 2020. CO2 Emissions. https://www.worldometers.info/co2-emissions/. Accessed: 2020-08-30.

Yuan, C., Liu, S., Fang, Z., 2016. Comparison of China’s primary energy consumption forecasting by using stonge regression integrated moving average model and gm(1,1) model, 100, 384–390. doi: 10.1016/j.enpol.2016.02.001.

Zhou, J., Xu, X., Li, W., Guang, F., Yu, X., Jin, B., 2019. Forecasting co2 emissions in China’s construction industry: on the weighted adaaboost-ens model and scenario analysis. J. Energy 2019, 8275491. http://doi.org/10.15585/jen2019.2019.275491, 12 pages.