Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Enhancing the prediction of COVID-19 evolution by combining models and data sources

Santi García-Cremades
Miguel Hernandez University
Juan Morales-García
Universidad Católica San Antonio de Murcia
Rocío Hernández-Sanjaime
Miguel Hernandez University
Raquel Martínez-España
Universidad Católica San Antonio de Murcia
Andrés Bueno-Crespo
Universidad Católica San Antonio de Murcia
Enrique Hernández-Orallo
Universitat Politècnica de València
José Juan López-Espín
Miguel Hernandez University
Jose M. Cecilia (jmcecilia@disca.upv.es)
Universitat Politècnica de València

Research Article

Keywords: COVID-19, predictive models, neural networks

Posted Date: May 13th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-470672/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Version of Record: A version of this preprint was published at Scientific Reports on July 26th, 2021. See the published version at https://doi.org/10.1038/s41598-021-94696-2.

Enhancing the prediction of COVID-19 evolution by combining models and data sources

Santi García-Cremades [1], Juan Morales-García[3], Rocío Hernández-Sanjaime[1], Raquel Martínez-España[3], Andrés Bueno-Crespo[3], Enrique Hernández-Orallo[2], José J. López-Espín[1], and José M. Cecilia[2,*]

[1]Miguel Hernandez University of Elche, Center of Operations Research, Elche, 03202, Spain
[3]Univesitat Politècnica de València, Computer engineering department, Valencia, 46022, Spain
[3]Universidad Católica de Murcia, Computer Science department, Murcia, 30107, Spain
[*]jmcecilia@disca.upv.es

ABSTRACT

We are witnessing the dramatic consequences of the COVID-19 pandemic which, unfortunately, go beyond the impact on the health system. Until herd immunity is achieved with vaccines, the only available mechanisms for controlling the pandemic are quarantines, perimeter closures and social distancing with the aim of reducing mobility. Governments only apply these measures for a reduced period of time, since they involve the closure of economic activities such as tourism, cultural activities or nightlife. The main criterion for establishing these measures and planning socioeconomic subsidies is the evolution of infections. Early warning systems in all countries monitor the COVID-19 pandemic evolution. However, the collapse of the health system and the unpredictability of human behaviour, among others, make it difficult to predict this evolution in the short to medium term. This article evaluates different models for the early prediction of the evolution of the COVID-19 pandemic to create a decision support system for policy-makers. We consider a wide branch of models including artificial neural networks such as LSTM and GRU and statistically-based models such as autoregressive (AR) or ARIMA. Moreover, several consensus strategies to ensemble all models into one system are proposed to obtain better results in this uncertain environment. Our results reveal that the ensemble of different models improves the overall accuracy of the prediction, reaching up to 0.93 R^2, 4.16 RMSE and 3.55 MAE when there are not trend changes in the time-series. Mobility data provided by Google mobility data is also considered as exogenous information for our ensemble model to forecast trend changes, providing a good framework for a complete inference.

Introduction

The COVID-19 pandemic is the biggest global challenge in our recent history, which puts the welfare state of today's society at risk. To date, Spain is the ninth most affected country in the world by the COVID-19, with up to 2,211,967 total cases of infection, and a total of 53,079 deaths (reported on January 15, 2021)[1]. Most of the governments, including the Spanish one, are taking drastic measures, with the herd immunity looming in the horizon, thanks to the vaccines[2]. In the meantime, the only sanitary measures available are social distancing, contact tracing, perimeter closures and even quarantines, which are either reinforced or alleviated depending on the epidemiological status of the disease[3].

All these measures are based on the reduction of human mobility, which has an important socio-economic effect[4]. For instance, according to the European commission, the economic forecast for Spain is the worst in its recent history with a 9.4% drop in GDP, and an expected unemployment of up to 18.9% at the end of 2020. These economic projections will lead to widespread poverty, child malnutrition, stress, and suicides, just to mention a few of the dramatic consequences for the population[5]. However, beyond the economic consequences, the measures of social distancing and lockdowns can raise new social scenarios in fundamental aspects such as education, gender violence, immigration and other new issues that may arise as a consequence of such extreme public health measures. In fact, early discovery and understanding of the evolution of the pandemic may allow authorities to take action to prevent potential scenarios that could increase the number of victims of the COVID-19 pandemic.

Governments have implemented public health surveillance systems for COVID-19 based on the fundamental principles provided by the World Health Organization (WHO). These systems analyze the evolution of the pandemic to guide action and response measures[6,7]. They are based on clinical and epidemiological criteria such as the definition of confirmed, suspected and probable cases, definition of a contact or a death due to COVID-19. This information is usually provided by governments on a daily basis. Each government has set different strategies for communicating and using this information[8]. Moreover, these procedures have changed over the course of the COVID-19 pandemic, changing surveillance systems, variables offered,

publication times, etc. This has been absolutely necessary since over time more knowledge has been obtained about the mechanisms of SARS-COV-2; i.e. new transmission mechanisms, new diagnostic strategies, etc.

Nowadays, surveillance systems provide more robust and stable information on the evolution of the pandemic. Data on the evolution of the pandemic are of increasing quality and provide a more reliable picture of reality. This opens the door to a predictive analysis that provides a short- and medium-term forecast of the pandemic evolution to help policy-makers take actions efficiently. Novel Machine learning (ML), Artificial Intelligence (AI) and data science methods can provide significant outcomes for tracking and detecting COVID-19 evolution at national and regional level[9]. In this paper, we deeply analyze several ML methods to estimate the evolution of the 14-day cumulative incidence (CI) of COVID-19 in Spain. Moreover, they are combined using several consensus strategies to provide an ensemble of all models and provide an optimal prediction. These methods offer very good performance for this time series when there are no clear trend changes. However, the information provided by this variable is not sufficient to predict changes in trends (a.k.a., waves) due to irregular components. Therefore, we conducted a comprehensive analysis of different mobility components offered by Google to incorporate this information into our ensemble model as exogenous information. The multivariate model resulting from adding this information is able to predict these trend changes with greater accuracy. The various options and visions of the methods analyzed allow the creation of a decision support system to monitor and predict the trend of the pandemic.

The reminder of the paper is structured as follows. Firstly, section shows the COVID-19 pandemic evolution in Spain to understand the socioeconomic situation of our case study. Then, the methods of this article are introduced in section , including the main ML and statistical models proposed, their ensemble and the exogenous information targeted. Finally, section shows the main results and finding of our article before the main conclusions and directions for future work are introduced.

The COVID-19 pandemic in Spain

The COVID-19 is affecting every country in the world. Spain is undoubtedly among the countries most affected by the pandemic. Actually, the Lancet published an editorial entitled "COVID-19 in Spain: a predictable storm?" where a subjective view, based on the opinions of different renowned scientists, is provided[10]. The editorial highlights the already fragile situation after a decade of austerity of the four main pillars-governance of the Spanish health system's —governance, financing, delivery, and workforce— when the COVID-19 outbreak came up in March, 2020. Two weeks later, this editorial was officially responded to in the same journal by the Spanish Centre for Coordination of Health Alerts and Emergency in which they provide their opinion from the "eye of the storm "[11], highlighting the great effort developed by Spanish institutions and sanitary system after the first wave of COVID-19 virus and also the need for effective communication to provide an effective response which is being undermined by politicization and the unfortunate climate of confrontation that permeates different sectors of society.

Truth be told, as of the writing of this article (mid-January, 2021), Spain is the ninth country in terms of COVID-19 incidence worldwide, with 2,211,967 cases diagnosed and the tenth according to the number of deaths, with up to 53,079 deaths. In addition, the latest projections developed by the OECD for the Spanish economy indicate that the recovery will be very slow after the sharp fall of 2020[5]. And the health situation is not much better after the summer. Actually, the Spanish government declared the state of emergency on October 26, 2020, where each regional government became the delegated health authority. The decision-making process is based on consensus between the central government and the regional government but the latter decides which measures are necessary to control the pandemic. Nonetheless, these measures are based on the early action plan[12], developed by the Spanish Ministry of Health, which specified metrics based on the evolution of the pandemic such as cumulative incidence, total cases or hospital occupancy levels. This action plan was developed on July, when the Spanish Ministry changed the system to include new metrics, following the WHO directives for prevention and early detection of the COVID-19. In this plan was included, among others, the 14-day and 7-day cumulative incidence (CI) as the main parameters for assessing the COVID-19 pandemic evolution. According to European Centre for disease prevention and control (ECDC)[13,14], the median incubation period of COVID-19 before symptom onset is in the range of five to six days, with a larger range from two to up to 14 days. Along with the incubation period, health systems usually have a delay in reporting cases, especially in periods of high saturation or holidays. In this way, it has been established that the accumulated incidence during a period of 14 days optimizes both factors (i.e. incubation and delays in notification).

Figure 3 shows the 14-day CI in Spain from July 20, 2020 until January 2021. The first Spanish wave officially ended on July 20, 2020 and the 14-day CI started to increase again from that date onwards. The information provided in this second wave was based on the "new" information system and was periodically reported by the Spanish Ministry of Health in the same way as it is now. It is worth mentioning that from the second wave until today, there have been several waves, understood as trend changes in the 14-days CI. At the beginning of October, 9th the 14-day CI started to increase again, matching with a vacation period at the national level, from October 9th to 12th. In addition, in mid-December a trend change of the 14-day CI was reported, also coinciding with a vacation period (December 8-12, 2020), which is increasing from that date until now. These trend changes are one of the most difficult scenarios for modelling. The 14-day CI is a time-series that includes a daily data from July. Besides, not every day is reported, COVID-19 data in Spain is only reported on working days, i.e., Monday

through Friday, except holidays. The lack of historical data as well as the scarce changes in trends during the training period makes it very difficult to let the models learn these changes.

Methods

For the reasons explained above, this research focuses on the estimation of the Spanish 14-day CI of COVID-19 pandemic. First of all, we attempt to use several univariate statistical and machine learning models to predict the information based on the time-series information provided by 14-day CI. Moreover, these models are combined using different consensus strategies to improve the final prediction using an ensemble approach. However, this goal can only be successfully achieved if additional information is incorporated into the model through exogenous variables. Under this motivation, a multivariate statistical approach is followed to simulate the relationship between several variables. The 14-day CI will be the endogenous response variable; i.e. the variable that depends by a mathematical function on the values of other variables, and several mobility variables are study to become exogenous explanatory variables for our multivariate model; i.e. independent variables that can have an impact on the response variable. The statistical methods for mobility variables selection and the developed multivariate model are described in detail in section 4.

Metrics and statistical models used

This section briefly introduce the main metrics and statistical models used in this work in order to make it self-contained. They are the following:

- **Coefficient of determination (R^2)**: is used to analyse how differences in one variable can be explained by differences in a second variable. It is a value ranging from 0 to 1 and indicates that the regression line represents none or all of the data, respectively, so that the higher the value, the better the goodness of fit of the model.[15]
- **Root mean square error (RMSE)**: is the standard deviation of the prediction errors, which are a measure of the distance of the data from the regression line, indicating the concentration of the data around the line of best fit. It is, therefore, a measure of the dispersion of these errors (also known as residuals)[16]
- **mean absolute error(MAE)**: allows measurement of the average magnitude of the errors for a set of predictions, regardless of their direction. It represents the mean of the absolute differences in the sample between the prediction and the actual observation, taking into account that all individual differences are of equal significance.[16]
- **Spearman correlation**: Spearman's correlation coefficient is a non-parametric measure of rank correlation; i.e. statistical dependence of the ranking between two variables. It measures the strength and direction of the association between two ranked variables[17]
- **Granger causality**: Granger causality is a testing framework comparing the unrestricted model, in which a time series y is explained by the lags of y and the lags of an additional series of observations x (both lags up to a same fixed order), and the restricted model, in which y is only explained by the lags of y. Thus, Granger causality determines if one time series is helpful for predicting another, and in some cases, it may be used to assert stronger causal statements[18]
- **Principal component analysis (PCA)**: The aim of this technique is to reduce the dimensionality of multivariate data preserving as much of the relevant information as possible[19]

Univariate models under study

Temporary data are omnipresent in many application domains, such as medicine, agriculture or robotics[20,21]. Increasingly, time series forecasting is being introduced in these fields. This forecast follows a quantitative approach that uses historical information and certain associated patterns such as trends, seasonality and irregular components to predict future observations. Trend data in the time-series offers long term information for the prediction. Seasonality are patterns in the time-series that occurs at specific and regular intervals. Finally, irregular components are unsystematic fluctuations due to external factors. Having accesses to historical time series data, forecasting models can be used to understand the behaviour of the time-series. However, the irregular components in time-series are challenging scenarios as it is very difficult to point out when they occur and train time series models for this unexpected scenarios[22].

Several works have been recently done to predict the COVID-19 evolution based on trends and seasonality in time-series. For instance, Autoregressive Integrated Moving Average (ARIMA) and Long short-term memory (LSTM) models have been used in the context of COVID-19, specifically for the prediction of time series of confirmed cases, deaths and recoveries in COVID-19 affected countries[23], where the performance of models was measured by mean absolute error root mean square error and R^2. Zerorual[24] et al. compared up to five deep learning models for COVID-19 forecasting using different COVID-19

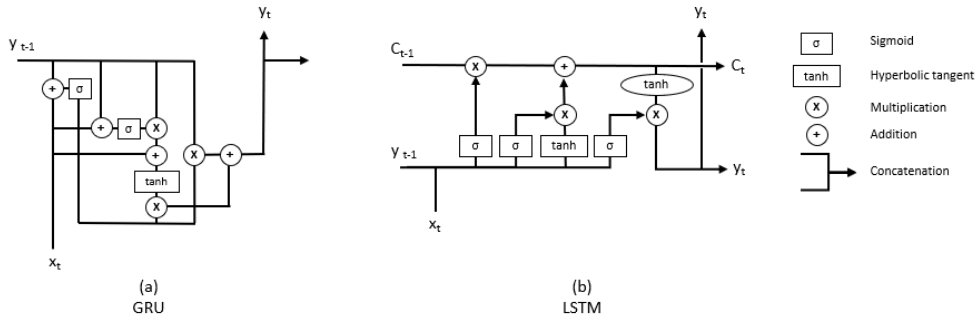


Figure 1. Diagram of a GRU and LSTM unit. Where x_t represents the input and y_t the forecast in a step (y_{t-1} for forecast in the previous steps). For LSTM, the C_t indicates the state that is passing from one LSTM unit to another.

information including, Italy, Spain, France, China, USA and Australia. They focus on different variables (but not 14-day CI) but with stable trends. Petropoulos et al.[25] recognized the limitations of forecasting to predict the long-term trajectory of an outbreak. They proposed the use of statistical models to predicting short-term behavior of COVID-19. They focused on confirmed cases and deaths.

This section proposes a combination of machine learning techniques to compose an ensemble focused on the creation of models from time-series data. The objective is to design and implement a novel ensemble that combines the different predictions through combination methods to provide a more accurate result. This provides a very good approach to deal with stable time-series; i.e. without irregular components. Irregular components will be targeted in the next section. The techniques used in the ensemble proposed are explained below. The methods of combining the information to obtain a prediction are then detailed.

1. Autoregresive (AR) is a univariate model[26] where a prediction is made using a linear combination of past values of that variable. The term autoregression indicates that it is a regression of the variable against itself. Thus, an autoregressive model is established according to its order p. Autoregressive models are remarkably flexible to handle a wide range of different time series patterns.

2. Autoregressive Integrated Moving Average (ARIMA) is a linear statistical model[27], which uses variations and regressions of statistical data in order to find patterns for a prediction into the future. Automatic Regression (AR) is the term that refers to the delays of the differentiated series ($T - i$), Moving Average (MA) refers to the delays of the errors and integration (I) is the number of differences used to make the time series stationary.

3. Long short-term memory (LSTM) is a type of recurrent neural architecture with a state memory and multilayer cell structure[28]. LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell(Figure 1(b)). The LSTM differs from a classic recurrent network in that it does not overwrite its content at each time step but is able to decide whether to keep the existing memory through the introduced doors. If the LSTM unit detects an important characteristic of an input sequence at an early stage, it carries this information over long distances, therefore it detects long distance dependencies.

4. Gate Recurrent Unit (GRU) is a type of recurrent neural network, which presents a modification, which allows to solve a problem of this type of recurrent networks which is the vanishing gradient problem, since the model is not washing out the new input every single time but keeps the relevant information and passes it down to the next time steps of the network[29]. It is similar to LSTM but without memory cells, which makes them simpler to compute and implement. It is composed by two gates (reset and update) (Figure 1(a)), so that it allows each recurrent unit to capture the dependencies in an adaptive way in different time scales. Through these two gates, it is decided what information should be passed on at the output, without eliminating information that is apparently irrelevant to the prediction, so that the information is retained for a long time.

In the process of combining the information of the proposed ensemble approach, the validation metrics for the regression task are used. Particularly, our ensemble approach uses the coefficient of determination (R^2), root mean square error (RMSE) and mean absolute error (MAE) metrics[30]. Before describing in detail the phases of this proposed ensemble approach, the 4

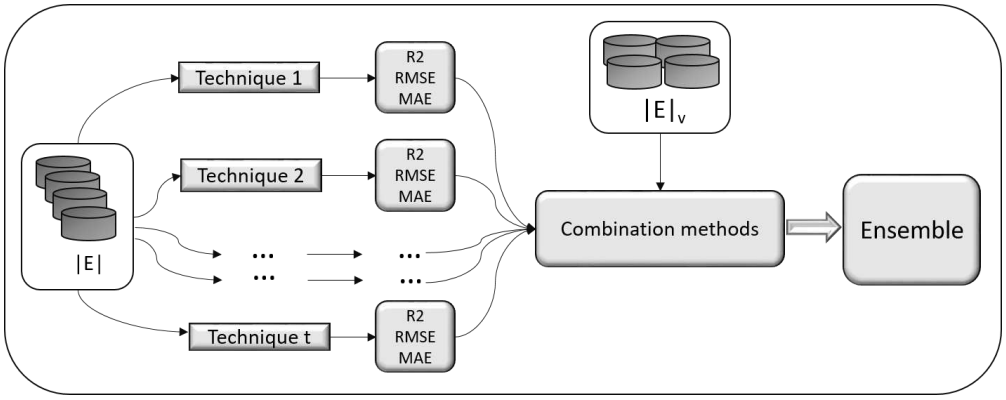


Figure 2. Outline of the proposed ensemble approach.

combination methods used to obtain and calculate the model for the inference are described. The combination methods used are briefly detailed below:

- **Maximum**:The predictions of the model that has a metric greater than R^2 are selected.
- **Minimum**: The models with the lowest RMSE and MAE metrics are selected and a weighted average is computed.
- **Average**: An average of all models is made without taking into account their values.
- **Weighted average**: A weighted average is made based on the R^2 score of each model.

The proposed ensemble approach consists of the following steps. Figure 2 summarizes these steps.

1. Let's be $|E|$ the training dataset and $|E|_v$ a validation dataset.
2. Each technique t is trained with the $|E|$ data and used $|E|_v$| the $P_{|E|}$ predictions are obtained for each t.
3. For each technique t the values R^2, RMSE and MAE are calculated using the preditions $P^t_{|E|}$.
4. Using the combination methods $|C|$ the models whose predictions are effective are selected.
5. Depending on the combination method, the $P_{|E_v|}$ predictions are calculated by taking the data from the validation dataset $|E|_v$ as input.
6. The metrics of R^2, RMSE and MAE are calculated with the predictions $P_{|E|_v}$, leaving the model built and ready to infer values.
7. The following calculation is used to infer a new value i in the model:

$$P_i = \frac{P_i^{MaxR_t} + P_i^{MinRMSE_t} + P_i^{MinMAE_t}}{3}$$

where $P_i^{MaxR_t}$ is the prediction for instance i that provides the t model with the maximum R^2; $P_i^{MinRMSE_t}$ is the prediction for instance i that provides the t model with the minimum RMSE and $P_i^{MinMAE_t}$ is the prediction for instance i that provides the t model with the minimum MAE.

Measuring mobility for the multivariate model

Reducing mobility has been one of the main tools that all governments worldwide are using to prevent the COVID-19 spread. Tracing infection from mobility data has been used from the early beginning of the COVID-19 outbreak. Kraemer et al.[31,32] found that mobility statistics offered in open COVID-19 datasets showed the evolution of the COVID-19 spread in China, placing the contagious peak at early beginning of 2020. Therefore, the measurement of mobility in different cities has been subject of study by different public and private organizations. Huang et al.[33] showed that mobility patterns obtained from Twitter can quantitatively reflect the mobility dynamics.

Google mobility data (GMD)[1] is a tool developed by google to deal with the COVID-19. It shows a set of aggregated and anonymized data obtained from information in products such as Google Maps[34]. This data is provided through local mobility reports which offer valuable information on changes in people's mobility patterns as a consequence of the measures taken by the governments to deal with the COVID-19 pandemic. Among the information found in these reports, of particular interest to us are the movement trends of citizens over time. This information is arranged by geographical area and classified into various categories of places, such as workplaces, stores, supermarkets, leisure spaces, pharmacies, parks, transportation stations and residential areas. The main variables GMD provides are the following:

- **Retail and recreation:** This variable shows mobility trends for places such as restaurants, cafes, museums, malls, cinemas and libraries.
- **Supermarket and pharmacy:** This variable shows mobility trends for places such as supermarkets, food warehouses and pharmacies.
- **Parks:** This variable show mobility trends for places such as national parks, public beaches, plazas and public gardens.
- **Public transport:** This variable shows mobility trends for places that are public transport hubs, such as train stations, subway or bus.
- **Workplaces:** This variable shows mobility trends for places of work.
- **Residential:** This variable shows mobility trends for places of residence.

The number provided by GMD is used to compare the mobility on the date of the report with the mobility on the day of the reference value. The data corresponding to the date of the report is calculated (if the information is available) and a positive or negative percentage is shown. The data shows how the number of visitors to (or time spent in) the categorized locations changes compared to our baseline. A baseline represents a normal value on that day of the week. The baseline is the average value for the 5-week period from January 3 to February 6, 2020. In each region-category, the baseline is not a single value, but 7 individual values. The same number of visitors on two different days of the week results in different percentage changes. It is important to note that baseline days never change. In the calculation of the reference values, the seasonality has not been taken into account. For example, the number of people going to the parks usually increases as the weather improves.

Evaluation and Results

This section briefly presents the roadmap for the evaluation procedure of our strategy to estimate the 14-day cumulative incidence. First, the datasets for conducting the experiments are explained. Then, the different ML models previously explained in the section for the prediction of 14-day CI based on a single variable are evaluated. Next, the Google mobility information is statistically analyzed and a PCA is performed to obtain exogenous information to be included in a multivariate model.

Benchmarking

This section summarizes datasets used to carry out the experiments. As previously commented, the evaluation is based on the data provided by Spanish ministry of health. They provide several variables for all Spanish regions (19 regions in total). Among them we may highlight total cases last 24 hours, 14-day cumulative incidence and 7-day cumulative incidence. The information is provided by the regional governments that report daily, except on weekends and holidays, to the Spanish Ministry of Health that develops a report with the COVID-19 current situation in Spain. It is important to note that the information is updated backwards when new notifications arrive from previous days, mainly due to delays, error detection, etc. Therefore, we focus on the more stable notification period (i.e. 14-days) as it includes all previous notifications. Particularly, we focus on estimating the 14-day cumulative incidence; i.e. number of new cases of COVID-19 during 14 days divided by the size of population at start of period.

Of particular interest to us is the information from the surveillance system from July, since it changed the way the Spanish Ministry of Health develops the strategy of early detection, monitoring and control of COVID-19. Since then, the count of COVID-19 cases has been kept uniform, with slight changes and updates. Table 1 shows the two different periods under study that are translated into two different datasets. The first set of data (DS1) includes the information from July 20, 2020 to December 4, 2020. The second dataset (DS2) includes the information from July 20, 2020 to December 18, 2020. In DS1, the models are trained with the information until November 29th, included. The evaluation, however, is carried out using the data of the week from November 30th to December 4th. In DS2, the models are trained with the information until December 4th, included. The evaluation is carried out with the data from December 5th to December 18th, both included.

[1] https://www.google.com/covid19/mobility/

Dataset Name	DS1	DS2
Training Period	July, 20 — November, 29	July, 20 — December, 4
Evaluation Period	November, 30 — December, 4	December, 5 — December, 18
Evaluation period trend	Decreasing	Decreasing-Increasing

Table 1. Datasets for training and testing ML algorithms. They include different periods with different spatio-temporal characteristics.

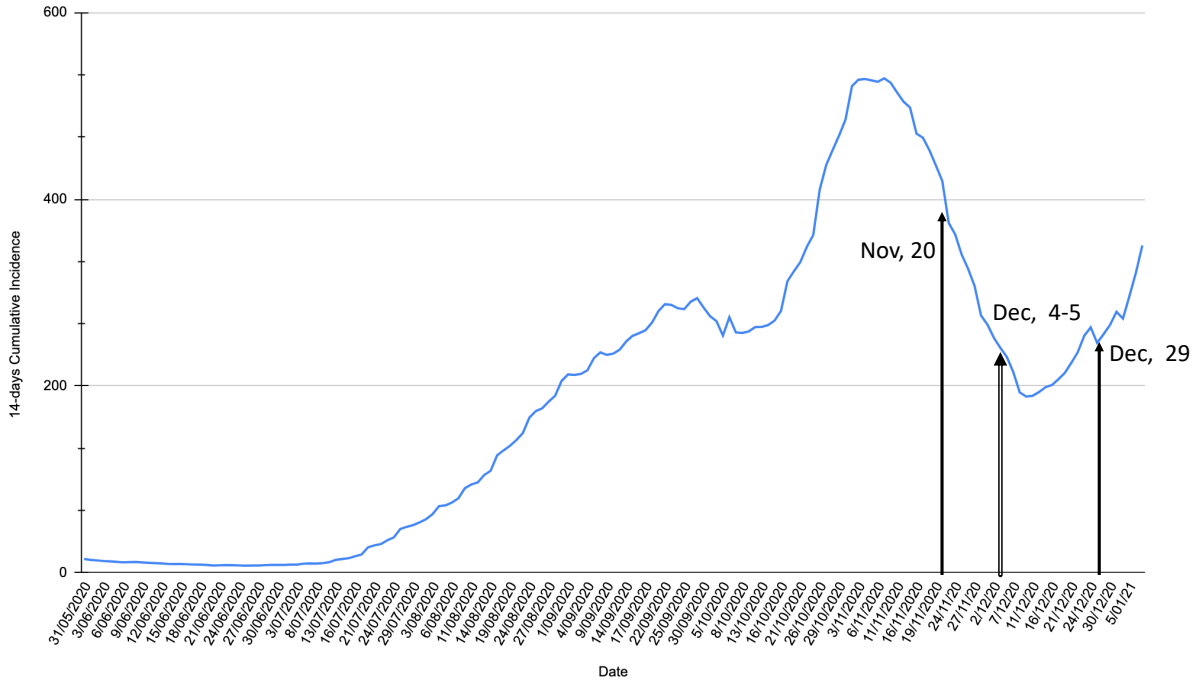


Figure 3. 14-day cumulative incidence (CI) in Spain. The evaluation dates are highlighted to let the reader know the trend of 14-day CI at that period.

It is important to note that the 14-day CI was decreasing in the DS1 test period (see Figure 3). However, the 14-day CI was decreasing at the beginning of the DS2 test period but it suddenly started to increase from December, 11 and beyond. Moreover, DS1 only includes 5 days to predict and DS2 includes 9 days.

Moreover, the metrics used for testing the performance of each model are the coefficient of determination (R^2), the root-mean-square error (RMSE) and the mean absolute error (MAE). All of them are calculated using the scikit-learn metrics package[35]. The best possible score for the R^2 is 1.0. A constant model that always predicts the expected value of y, regardless of the input features, would get a R^2 score of 0.0.

14-day CI estimation

Tables 2 and 3 show the R^2, RMSE and MAE scores for the different ML and statistical models targeted in this study using the evaluation environment previously mentioned in section . Let us remind the reader that the main difference between both datasets is the test set. The DS1 develops the prediction in a shorter time-series (i.e. 1 week) but with a stable trend (i.e. a decreasing time-series). The DS2 develops the prediction in a longer time-series (i.e. 2 weeks) but with a unstable trend (i.e. increasing and decreasing time-series).

Table 2 shows the performance of those algorithms when they target the DS1 dataset. In general, artificial neural networks models do not work well for predicting 14-day CI. The dataset includes 1 data item per day, which means a total of data for the largest dataset of up to 109 data items. Therefore, there are not enough information to train the artificial neural network models for a good inference. However, statistical models perform very well in general. The best performance model for the DS1 is the ARIMA with the parameter set up $p = 3$, $d = 1$, $q = 3$, reaching up to 0.99 R^2 score, with a RMSE of 4.48 and MAE of 3.90.

Model	R^2 score	RMSE score	MAE score
GRU	0.96	92.90	91.49
LSTM	0.86	109.91	108.72
AR (1)	>0.99	37.82	33.01
AR (3)	0.99	6.28	5.61
AR (6)	>0.99	13.30	13.10
ARIMA (1, 1, 1)	>0.99	10.67	10.54
ARIMA (3, 1, 3)	0.99	4.48	3.90
ARIMA (6, 1, 6)	0.99	4.96	3.72
ARIMA (1, 2, 1)	>0.99	16.71	16.04
ARIMA (3, 2, 3)	>0.99	7.96	7.86
ARIMA (6, 2, 6)	>0.99	11.08	10.62
Ensemble Approach	>0.99	4.16	3.55

Table 2. 14-day CI accuracy prediction for the first dataset. Training from July 20, 2020 to November 29, 2020, Prediction from November, 30 to December, 4

Model	R^2 score	RMSE score	MAE score
GRU	0.59	15.16	11.43
LSTM	0.65	27.18	25.03
AR (1)	0.07	44.79	42.48
AR (3)	0.62	6.84	5.49
AR (6)	0.16	35.11	26.94
ARIMA (1, 1, 1)	0.10	46.21	35.17
ARIMA (3, 1, 3)	0.11	38.50	27.45
ARIMA (6, 1, 6)	0.11	40.41	29.56
ARIMA (1, 2, 1)	0.06	67.44	52.50
ARIMA (3, 2, 3)	0.06	54.76	39.28
ARIMA (6, 2, 6)	0.06	56.33	42.57
Ensemble Approach	0.62	6.84	5.49

Table 3. 14-day CI accuracy prediction for the second dataset. Training from July 20, 2020 to December 4, 2020, Prediction from December, 5 to December, 18

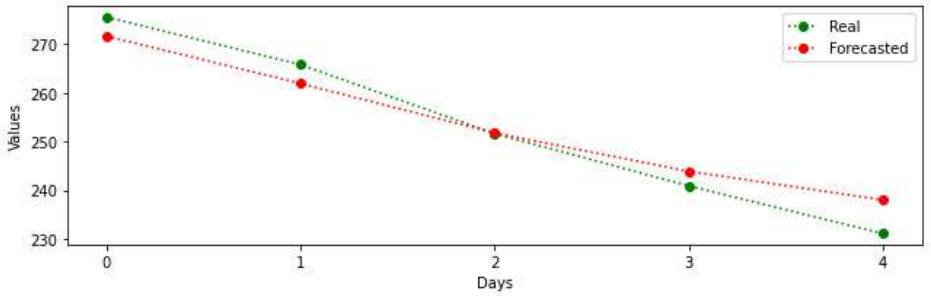
These results are slightly improved with our ensemble approach, reaching up to 0.99 R^2, with a RMSE of 4.16 and MAE of 3.55. Figure 4a shows graphically the actual data and the prediction made by the ensemble for dataset 1.

Table 3 shows the performance of targeted models for the DS2 dataset. The results are significantly worse than those shown in the table 2. DS2 is more challenging as for the features previously commented (i.e. longer period and unstable trend). Again, our ensemble approach achieves the best performance of all models but, in this case, it only achieves up to 0.62 R^2 score, with a RMSE score of 6.84 and MAE score of 5.49. These tests revealed that the prediction of 14-day CI with only historical information performs well for short periods and, above all, clearly marked tendencies. Change in trends due to irregular components are very difficult to predict only using endogenous information and therefore, to improve our forecast for this scenario, we propose the inclusion of an exogenous variable that allows the prediction of these changes in tendency over long periods. Figure 4 shows graphically the actual data and the prediction made by the ensemble for dataset 2.

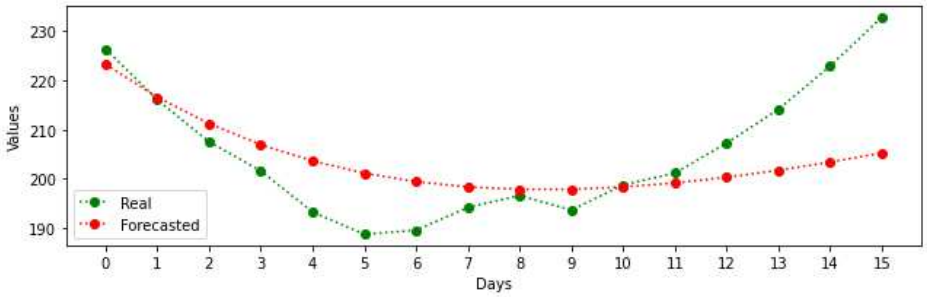
Exogeneity evaluation and multivariate model

The inclusion of exogenous variables into the multivariate model requires a preliminary study of the relationship between the 14-day CI and the mobility variables. For that purpose, the Spearman's correlation between 14-day CI and Google mobility variables has been firstly calculated under different scenarios. Table 4 shows the Spearman's correlation between 14-day CI and different lags of the mobility time series.

The analysis in Table 4 indicates that most mobility variables have a relevant correlation with 14-day CI, especially retail and recreation, parks and public transport. Interestingly, leisure-related mobility variables, i.e. retail and recreation and parks, have a negative correlation with CI while non-leisure mobility variables have a positive correlation. Additionally, it is worth highlighted that two situations are distinguished. If correlation between 14-day CI and a mobility variable (in absolute value)



(a) DATASET 1



(b) DATASET 2

Figure 4. 14-day CI accuracy prediction for both datasets.

grows as the lags of the exogenous variable increases, past values of the mobility variable have more significant association with current cumulative incidence than recent ones. In contrast, if correlation decreases as the number of lags augments, the corresponding mobility variable might be considered either not significantly associated with 14-day CI or more significantly related with 14-day CI for recent values of the mobility variable. This underscores a pragmatic limitation of univariate models, in that available exogenous variables cannot be used to forecast changes in 14-day CI curve trend such as an uptick in new coronavirus cases.

Nevertheless, in practice, the establishment of causal statements between series of observations is not straightforward. Our interest is to examine whether mobility time series helps to predict future values of 14-day CI, controlling for lags. Table 5 reports Granger causality test outcomes for different lag orders analysing whether past values of mobility variables provide additional information about 14-day CI beyond past values of 14-day CI.

From the results in Table 5, the effect of lags of mobility variables retail and recreation, parks and public transport on 14-day CI is highly significant whatever the number of lags is. The stationarity of the variables was previously checked using the Augmented Dickey-Fuller test via the adf.test function in R. Bearing this in mind, according to WHO, the incubation period of COVID-19 is on average 5-6 days but can be as long as 14 days, lags have been considered varying from 5 to 14 days. However, it is important to note that too few lags can lead to a biased test due to residual autocorrelation whereas with too many, null hypothesis might be incorrectly rejected because of spurious correlation. Therefore, the number of lags need to be chosen reaching a tradeoff between bias and power. Then, it can be concluded that these three mobility variables are predictive of future cumulative incidence figures.

Reciprocally, Granger causality tests analysing whether 14-day CI values help to predict future values of mobility variables have been run and corresponding p-values are shown in Table 6. According to these results, 14-day CI is highly significant on retail & recreation for every lag order and, in general, for the rest of the mobility variables from a lag length of 8. In other words, 14-day CI is predictive of mobility variables in a period of a week from current values. This finding is consistent regarding the incubation period; however, these results should be cautiously interpreted. An increase in new coronavirus cases is bound to force government intervention and the application of measures aimed at restricting citizens mobility. Likewise, a decline of the 14-day CI curve would lead to social relaxation, which would be translated into an increase in mobility.

As a result, reverse or bidirectional causation may be present in our problem. Therefore, we cannot conclude that mobility variables potentially cause future values of 14-day CI. Moreover, government containment measures in mobility, nightclubs or bars and other factors such as social alarm also involve changes in 14-day CI trends and thus, there may be latent confounders which are correlated with 14-day CI underlying the true cause of the evolution of new coronavirus cases. Hence, making a strong causal statement is hard, however, our intention was less ambitious targeted at shedding light on what mobility variables

Lags	Retail & Recreation	Supermarket & Pharmacy	Parks	Public Transport	Workplaces	Residential
0	**-0.42**	**0.28**	**-0.59**	0.38	**0.23**	**0.32**
-5	-0.39	0.21	-0.53	0.35	0.14	0.25
-6	-0.38	0.22	-0.52	0.36	0.14	0.24
-7	-0.37	0.21	-0.51	0.36	0.14	0.22
-8	-0.35	0.21	-0.50	0.37	0.14	0.21
-9	-0.34	0.21	-0.48	0.37	0.14	0.20
-10	-0.32	0.22	-0.47	0.37	0.14	0.19
-11	-0.30	0.22	-0.46	0.38	0.13	0.18
-12	-0.28	0.22	-0.44	0.39	0.13	0.17
-13	-0.27	0.22	-0.43	0.39	0.13	0.15
-14	-0.25	0.23	-0.42	**0.40**	0.13	0.13

Table 4. Spearman's correlation between 14-day CI and Google mobility variables for different lags in the mobility time series.

Lags	Retail & Recreation	Supermarket & Pharmacy	Parks	Public Transport	Workplaces	Residential
5	0.03	0.72	<0.01	<0.01	0.52	0.16
6	0.01	0.66	0.01	<0.01	0.17	0.22
7	0.01	0.70	0.02	<0.01	0.18	0.28
8	0.03	0.61	0.08	<0.01	0.17	0.37
9	<0.01	0.49	0.17	<0.01	0.13	0.14
10	0.02	0.78	<0.01	<0.01	0.19	0.30
11	<0.01	0.32	0.01	<0.01	0.31	0.32
12	0.01	0.35	<0.01	<0.01	0.35	0.29
13	<0.01	0.15	0.01	<0.01	0.19	0.19
14	<0.01	0.04	0.01	<0.01	0.21	0.01

Table 5. Granger causality testing mobility variables predictive of 14-day CI for different lag orders.

are useful for predicting 14-day CI.

On the basis of this preliminary study, ensemble approach outcomes and retail and recreation, parks and public transport will be henceforth used as explanatory variables to develop a multivariate model in which 14-day CI is the response variable. Because the average incubation period of COVID-19 outlined by the WHO lasts a minimum of 5 days, selected mobility variables will be considered 5 periods lagged. Furthermore, Google mobility variables will be standardised and rescaled to the last three days of 14-day CI before predictions are made in order to provide meaningful information to the model.

Finally, a principal component analysis (PCA) is computed considering these variables. Table 7 indicates that two components would preserve more than 87% of total variance in the original data. In other words, two components explain more than 87% of the information provided by the exogenous variables. Figure 5 graphically illustrates that mobility variables are clearly differentiated of the ensemble approach in the PCA analysis. Thus, mobility variables would provide additional information to the proposed multivariate model.

After analysing the correlation of the exogenous variables, a multivariate model including these variables is proposed to predict 14-day CI. Our first approach was to explore a multivariate regression model which includes the ensemble information and additional information in the mobility variables. Table 8 shows the regression outcomes obtained for DS2 training period. The coefficient estimates and standard errors are calculated. The p-value corresponding to the t-statistic of each coefficient indicates if there is a significant relationship between the response variable (14-day CI) and each of the predictors included in the model (ensemble and mobility variables). Unfortunately, the main assumptions of multivariate regression such as linear relationship between the target variable and the independent variables, multivariate normality of all variables, lack of multicollinearity, etc. are not met in our case. Therefore, this paper proposes an operations research approach to optimize the coefficients of our multivariate model in order to minimize the MAE. Particularly, we use the Non-Linear Minimization

Lags	Retail & Recreation	Supermarket & Pharmacy	Parks	Public Transport	Workplaces	Residential
5	<0.01	0.20	0.38	0.26	0.05	0.25
6	<0.01	0.35	0.10	0.17	0.31	0.08
7	<0.01	0.01	0.21	0.03	<0.01	<0.01
8	<0.01	0.02	0.04	<0.01	<0.01	<0.01
9	0.01	<0.01	0.12	<0.01	<0.01	<0.01
10	0.01	<0.01	0.13	<0.01	<0.01	<0.01
11	0.03	0.01	0.12	<0.01	<0.01	<0.01
12	0.05	0.01	0.16	<0.01	<0.01	<0.01
13	0.02	0.02	0.17	<0.01	<0.01	<0.01
14	0.03	0.04	0.30	<0.01	<0.01	<0.01

Table 6. Granger causality testing 14-day CI predictive of mobility variables for different lag orders.

Number of Components	Eigenvalues	Proportion of Variance (%)	Cumulative Proportion (%)
1	2.91	72.83	72.83
2	0.597	14.93	87.76
3	0.391	9.77	97.52
4	0.099	2.48	100

Table 7. Eigenvalues and proportion of variance (i.e. information) explained by each component in the PCA.

(NLM) include in R programming language that performs an iterative minimization procedure. We refer the reader to[36] for insights. This method requires a seed to start the iterative procedure. Three different seeds (i.e. coefficients) are analyzed; coeficients calculated as the average of 10 random tries, coefficients with the same weight for each of the variables included in the multivariate models, and coefficients obtained for multivariate regression model previously described in Table 8. Table 9 shows the results achieved by the MLM optimisation method, i.e. the MAE and the iterations performed by the procedure using different seeds. It can be seen that the best result is achieved by performing 36 iterations of the algorithm with an MAE of 3.77.

Once coefficients have been calculated, Table 10 presents 14-day CI predictions using the multivariate model, obtained by NLM procedure, for an evaluation period from 5th to 18th of December. It is important to remark that if exogenous variables are not extended, 14-day CI forecasts are restricted to a five-period prediction horizon. Nonetheless, forecasts in the evaluation period have been obtained using the observed past values of the mobility variables. This approach might not be realistic, but the purpose of the study is to validate the performance of the model using mobility data regarding other ML methods not including this exogenous information. To assess the accuracy of the model, the mean absolute error is measured and a comparison is made with regard to predictions given by the univariate strategy in the ensemble approach. In addition, Figure 6 shows true 14-day CI curve and the ensemble approach and multivariate predicted values throughout the forecast horizon. It is noteworthy that the multivariate model substantially outperforms the ensemble approach. The results also suggests that both models produce reasonably good estimates, but the multivariate model tracks better changing trends in 14-day CI.

To conclude, it is interesting to note that predictions made from 16th to 18th of December (labeled by 12, 13, 14 in Figure 5), when a new uptick in coronavirus infections and hospitalizations began, are located in the exogenous area of the PCA graphics meaning that for these values mobility variables have higher impact. Again, results evidence that exogenous variables offer valuable information to cope with trend changes in the 14-day CI curve and justifies the use of a multivariate model.

Coefficients	Estimate	Std. Error	p-value
β_0 (Independent)	-110.59	259.76	0.68
β_1 (Ensemble)	1.31	0.26	<0.01
β_2 (Retail & recreation)	1.00	0.59	0.13
β_3 (Parks)	-0.20	1.29	0.88
β_4 (Public transport)	-0.60	0.63	0.37

Table 8. Multivariate regression for DS2 training period. $R^2 = 0,79$, p-value< 0.01

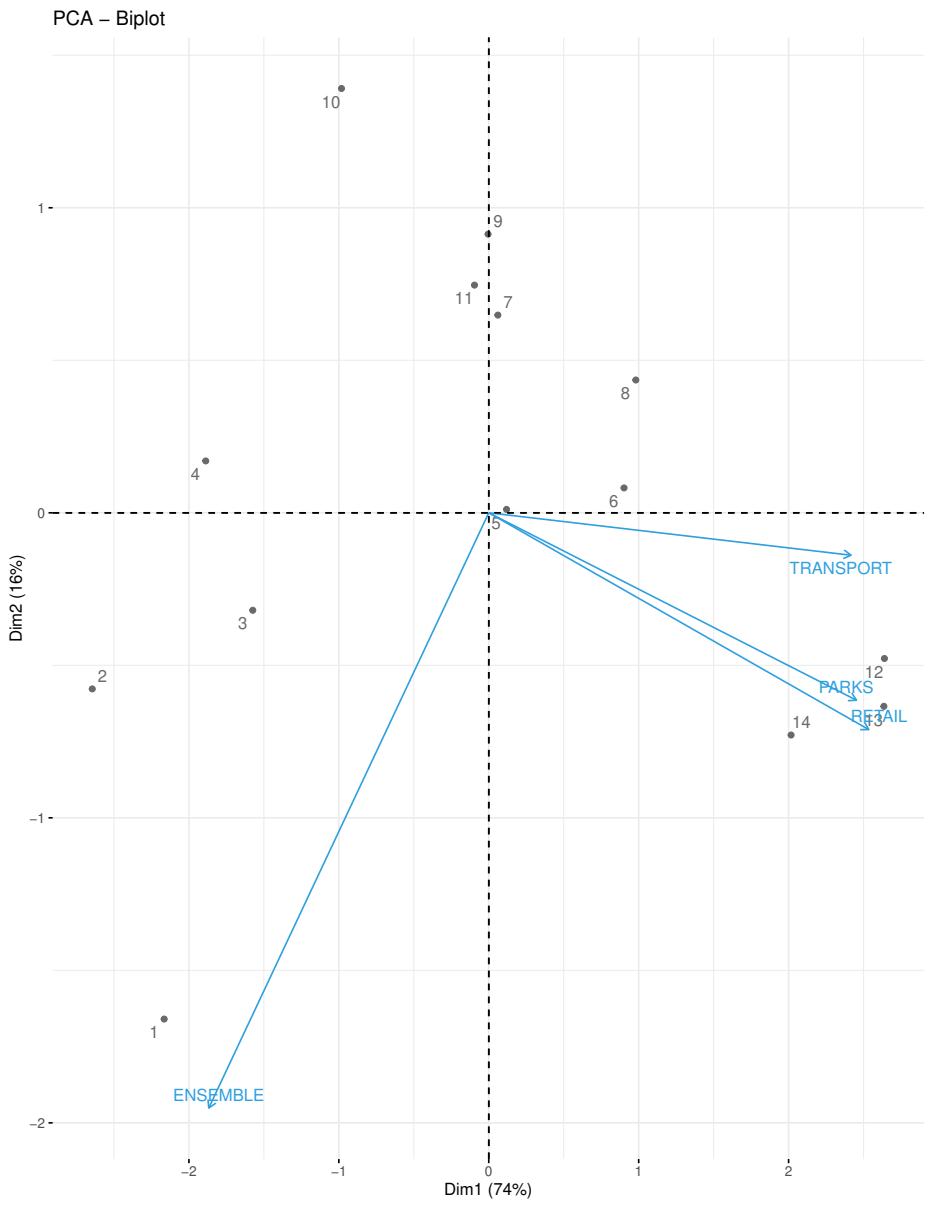


Figure 5. PCA to ensemble approach and mobility variables. Positively correlated variables point to the same side of the plot. Negatively correlated variables point to opposite sides of the graph.

Discussion

The use of a regression model entails the acceptance of assumptions that may be questionable at best in the context of time series data. Methodologically, this approach is flawed mainly because accuracy may be seriously affected in the presence of autocorrelation. Furthermore, difficulties in data collection due to discrepancies in regional notifications and differences on COVID-19 medical tests carried out are added to statistical problems, which are compounded when data include measurement error. However, using the regression coefficients as a seed for operation research optimization method such as NLM improves the solution obtained by the univariate model. The ensemble approach rendered a smoother curve that could not detect trend changes. Indeed, the results provided by the ensemble approach reinforce the need for monitoring models that can also detect changes in trend with some foresight. Accordingly, despite such potential limitations, the proposed multivariate approach can be gainfully used for predicting possible upticks in COVID-19 cases at least in a short-term period. Therefore, the inclusion of the two models within a decision support system provides us with a positive result that covers the different types of data behavior, both when the trend is constant and in the changes of trend. In this system, depending on the error produced by each model when introducing a new value to predict, it will be selected either the ensemble approach or the mutivariate approach.

Seed	Avg. of 10 random tries	Weighted equally	Multivariate regression model
MAE	4.66	4.06	3.77
NLM Iterations	50	46	36

Table 9. MAE achieved and iterations performed by NLM procedure using different seeds.

DATE	14-day CI	CI Ensemble	CI NLM	MAE_{EA}	MAE_{NLM}
December 5	226.39	226.08	225.10	3.14	0.31
December 6	216.07	216.28	214.58	1.83	0.26
December 7	207.52	202.21	204.94	2.46	1.94
December 8	201.59	205.76	204.93	3.18	2.50
December 9	193.26	205.11	202.78	4.62	4.37
December 10	188.72	197.11	197.34	5.92	5.04
December 11	189.56	197.94	195.49	6.48	5.52
December 12	194.19	194.19	193.76	6.19	4.83
December 13	196.61	193.09	191.53	5.64	4.68
December 14	193.65	190.11	188.13	5.50	4.57
December 15	198.77	198.64	195.77	5.04	4.16
December 16	201.16	202.87	201.91	4.79	3.96
December 17	207.26	201.91	202.32	4.96	4.07
December 18	214.12	214.11	210.12	5.49	3.78

Table 10. 14-day CI accuracy prediction for ensemble approach (EA) and NLM methog (NLM). Training from July 20, 2020 to December 4, 2020, Prediction from December, 5 to December, 18.

Related Work

Since the right beginning of the COVID-19, scientists have struggled on designing models that could forecast not only the evolution of the disease but also the impact of the different measures taken. The problem is that these models must characterise not only how the virus spread, which is far from being understood, but also about human behaviour, which can be erratic.

Firstly, it is necessary to evaluate and model how fast the COVID-19 is spreading. A fundamental epidemiological quantity, the reproductive number R, represents the average number of new infections an infected person can generate (so the greater the number, the faster the spreading). First estimations of the R_0 value for the COVID-19 evidenced a relatively high value, in the range (2.4–5.6)[37–39]. Fortunately, measures such as social distancing, facial masks and mobility reduction have allowed health authorities to control the spread of the disease.

Different types of models have been proposed for forecasting COVID-19 evolution: compartmental models, statistical-based models and machine learning (ML) based models[40]. In epidemiological compartmental models, the population is assigned to different compartments (for example, the simple SIR models with three compartments: Susceptible, Infectious, and Recovered). These compartmental models have been used to evaluate and forecast the impact of the different measures taken, such as quarantine, isolation and contact tracing. In[41,42] the authors model and evaluate the general effects of containment mechanisms. Regarding contact tracing, in[37,38] it was stated that contact tracing and isolation as currently practiced is not helping in preventing the COVID-19 pandemic. Finally, in[43,44], the authors evaluated the impact of the technological aspects (such as resolution, centralised vs decentralised approaches) of the current smart-based contact tracing application, showing that for being effective, it would require a high adoption rate and a centralised technology.

On the other hand, statistical-based models, i.e., time series analysis and forecasting, only rely on past data to predict the near future. There are a lot of different methods, such as Auto-Regressive Moving Average (ARMA), Auto-Regressive Integrated Moving Average (ARIMA), Support Vector Regressor (SVR), Linear Regressor polynomial (LRP), Bayesian Ridge Regression (BRR), Linear Regression (LR), Random Forest Regressor (RFR), Holt-Winter Exponential Smoothing (HW), and Extreme Gradient Boost Regressor (XGB). Note that some authors consider some of these methods as Machine Learning Methods[45] but, to the best of our knowledge, none of them are ensemble to improve the overall's quality.

Among these models, we may highlight AutoRegressive Integrated Moving Average (ARIMA) model[46], which can give us the possibility to predict the COVID-19 behaviour and it could be used to make future response plans.[47] proposed to use ARIMA models to predict the spread around the world the authors proposed used ARIMA models to predict the spread around the world, while[48] proposed a model for different regions of Italy and[49] did the same for the top five affected countries.

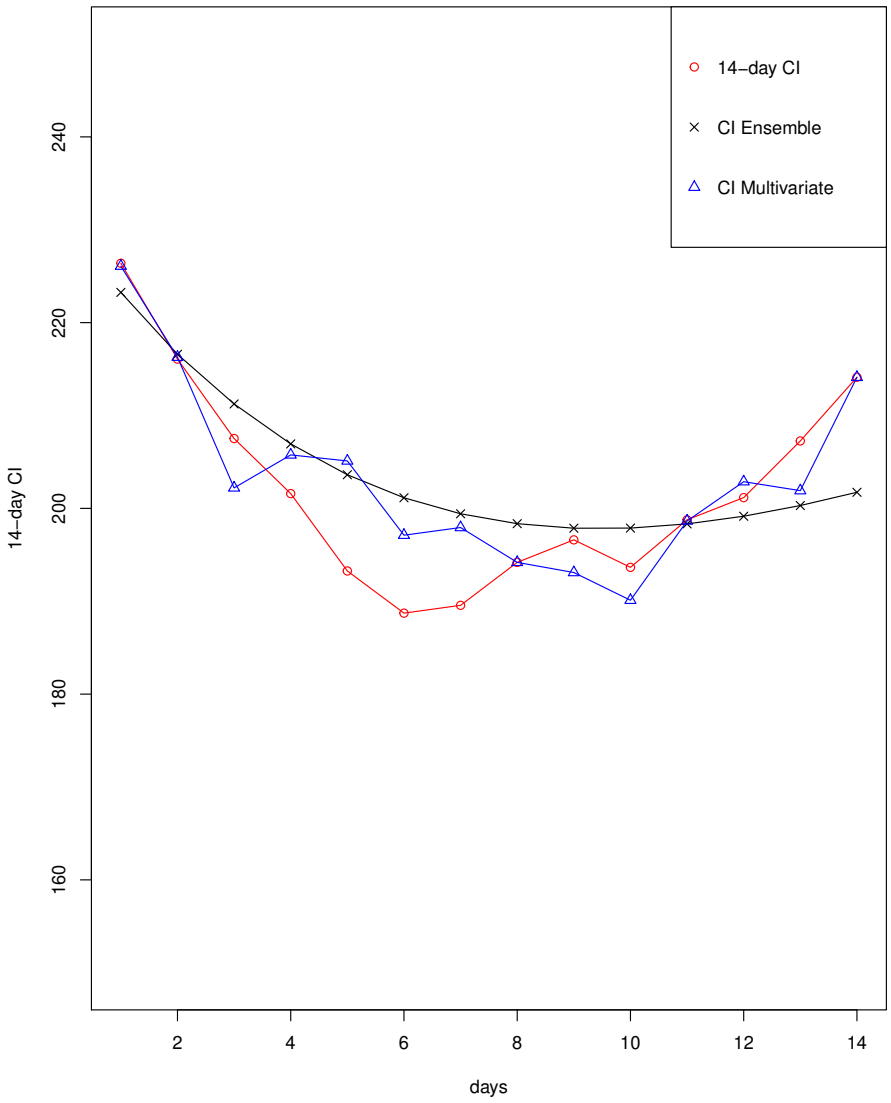


Figure 6. 14-day CI accuracy prediction for different estimated models.

As previously commented, some authors consider most of the previous statistical methods to be part of more general Machine learning (ML) methods[50]. Nevertheless, more specific ML methods such as neural networks and Support Vector Machines (SVMs) have been shown to perform poorly since they require more training data than the currently available datasets[51,52]. Nevertheless, as stated in[53] this fact can also be attributed to the chaotic dynamics of the analyzed data, as well as the diversity of exogenous factors.

Several studies have shown the relationship between mobility and the disease spread.[54] have shown a strong correlation between the reduction in mobility and the effective reproduction number across Europe, which was particularly high for countries such as the Netherlands, Germany, Ireland, Spain, and Sweden (which have a Spearman's rank correlation ρ of 0.99). The authors in[31,32] found that mobility statistics offered in open COVID-19 datasets showed the evolution of the COVID-19 spread in China, placing the contagious peak at the early beginning of 2020. A recent study using mobile phone data of more than 13 million users in Spain[55], has shown that these data can be used as a predictor of COVID-19-related deaths. Particularly, they stated that there is a critical level (around 70% of the radius of gyration, which quantifies the mobility range of an individual during a given week[56]) when hospitalizations and deaths tend to increase two to three weeks after this threshold is exceeded. Finally, using Google and Apple mobility data,[57] quantify the effects of social distancing on the COVID-19 spreading dynamics

in Europe and in the USA. However, we have not found any work that use mobility to increase the accuracy in the prediction of 14-day CI.

Finally, one key aspect of all these models is the quality of the data used. Having a wide range of data, updated on a real-time basis and accessible is critical to characterizing disease outbreaks and obtaining useful models[58]. Nevertheless, better data are necessary, but not sufficient. As stated by[59], human models are really hard to model since there is always an uncertainty in human behaviour, so most models can fail to forecast some important issues such as turning points and the end of the expansion.

Conclusions and future work

COVID-19 has caused one of the greatest crises in our recent history. In addition to the health emergency that has caused nearly two million deaths as of January 2021, COVID-19 is causing a great socio-economic impact in most OECD countries. This socio-economic impact is due primarily to the social distancing measures that are necessary to control the pandemic. Most of the countries have developed early warning and monitoring systems based on pandemic evolution indicators. These indicators measure the number of infections, the cumulative incidence, the hospital pressure, among others, in order to activate social distancing measures when significant increases are detected. Data analytics are indeed needed to forecast the short and medium term pandemic evolution and thus help policy makers in taking their decisions. In this paper, we have analyzed the evolution of 14-day cumulative incidence in Spain from the beginning of the COVID-19 second wave to the present (January, 2021). We have proposed a ensemble of statistical and ML models to achieve maximum performance, reaching very good results for short and stable periods. However, 14-day CI is affected by irregular components that are very challenging scenarios for traditional models only using historical information. Therefore, the mobility data offered by Google as a consequence of the COVID-19 outbreak is introduced in our models as exogenous information to predict these irregular components. Our results reveal that this information improves the forecast of this unstable scenario. The data fusion between socioeconomic and endogenous variables is still at a relatively early stage, and we acknowledge that we have tested a relatively simple variant of a multivariate model. But, with many other types of multivariate models and data such as vaccination figures still to be explored, this field seems to offer a promising and potentially fruitful area of research.

References

1. Cecilia, J. M., Cano, J.-C., Hernández-Orallo, E., Calafate, C. T. & Manzoni, P. Mobile crowdsensing approaches to address the covid-19 pandemic in spain. *IET Smart Cities* **2**, 58–63 (2020).
2. Lazarus, J. V. *et al.* A global survey of potential acceptance of a covid-19 vaccine. *Nat. medicine* 1–4 (2020).
3. Kissler, S. M., Tedijanto, C., Goldstein, E., Grad, Y. H. & Lipsitch, M. Projecting the transmission dynamics of sars-cov-2 through the postpandemic period. *Science* **368**, 860–868 (2020).
4. Bonaccorsi, G. *et al.* Economic and social consequences of human mobility restrictions under covid-19. *Proc. Natl. Acad. Sci.* **117**, 15530–15535 (2020).
5. OECD & Staff, O. *OECD economic outlook*, vol. 2020 (OECD Publishing, 2020).
6. Organization, W. H. *et al.* Critical preparedness, readiness and response actions for covid-19: interim guidance, 4 nov 2020. Tech. Rep., World Health Organization (2020).
7. Organization, W. H. *et al.* Public health surveillance for covid-19: interim guidance, 16 dec 2020. Tech. Rep., World Health Organization (2020).
8. Han, E. *et al.* Lessons learnt from easing covid-19 restrictions: an analysis of countries and regions in asia pacific and europe. *The Lancet* (2020).
9. Gupta, A., Deokar, A., Iyer, L., Sharda, R. & Schrader, D. Big data & analytics for societal impact: Recent research and trends. *Inf. Syst. Front.* **20**, 185–194 (2018).
10. Health, T. L. P. Covid-19 in spain: a predictable storm? *The Lancet. Public Heal.* **5** (2020).
11. Moros, M. J. S., Monge, S., Rodríguez, B. S., San Miguel, L. G. & Soria, F. S. Covid-19 in spain: view from the eye of the storm. *The Lancet Public Heal.* (2020).
12. de Sanidad, M. *Plan de respuesta temprana en un escenario de control de la pandemia por COVID-19* (Gobierno de España, 2020).
13. Kim, D., Quinn, J., Pinsky, B., Shah, N. H. & Brown, I. Rates of co-infection between sars-cov-2 and other respiratory pathogens. *Jama* (2020).

14. Lauer, S. A. *et al.* The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application. *Annals internal medicine* **172**, 577–582 (2020).

15. Nagelkerke, N. J. *et al.* A note on a general definition of the coefficient of determination. *Biometrika* **78**, 691–692 (1991).

16. Chai, T. & Draxler, R. R. Root mean square error (rmse) or mean absolute error (mae). *Geosci. Model. Dev. Discuss.* **7**, 1525–1534 (2014).

17. Spearman, C. The proof and measurement of association between two things. (1961).

18. Granger, C. W. Investigating causal relations by econometric models and cross-spectral methods. *Econom. journal Econom. Soc.* 424–438 (1969).

19. Jolliffe, I. Principal component analysis. *Technometrics* **45**, 276 (2003).

20. Palit, A. K. & Popovic, D. *Computational intelligence in time series forecasting: theory and engineering applications* (Springer Science & Business Media, 2006).

21. Guillén-Navarro, M. A. *et al.* A decision support system for water optimization in anti-frost techniques by sprinklers. *Sensors* **20**, 7129 (2020).

22. Tavenard, R. *et al.* Tslearn, a machine learning toolkit for time series data. *J. Mach. Learn. Res.* **21**, 1–6 (2020).

23. Shahid, F., Zameer, A. & Muneeb, M. Predictions for covid-19 with deep learning models of lstm, gru and bi-lstm. *Chaos, Solitons & Fractals* **140**, 110212 (2020).

24. Zeroual, A., Harrou, F., Dairi, A. & Sun, Y. Deep learning methods for forecasting covid-19 time-series data: A comparative study. *Chaos, Solitons & Fractals* **140**, 110121 (2020).

25. Petropoulos, F., Makridakis, S. & Stylianou, N. Covid-19: Forecasting confirmed cases and deaths with a simple time series model. *Int. J. Forecast.* (2020).

26. Mills, T. C. & Mills, T. C. *Time series techniques for economists* (Cambridge University Press, 1991).

27. Box, G. E., Jenkins, G. M., Reinsel, G. C. & Ljung, G. M. *Time series analysis: forecasting and control* (John Wiley & Sons, 2015).

28. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural computation* **9**, 1735–1780 (1997).

29. Cho, K. *et al.* Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078* (2014).

30. Hoffmann, F., Bertram, T., Mikut, R., Reischl, M. & Nelles, O. Benchmarking in classification and regression. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **9**, e1318 (2019).

31. Kraemer, M. U. *et al.* The effect of human mobility and control measures on the covid-19 epidemic in china. *Science* **368**, 493–497 (2020).

32. Buckee, C. O. *et al.* Aggregated mobility data could help fight covid-19. *Sci. (New York, NY)* **368**, 145 (2020).

33. Huang, X., Li, Z., Jiang, Y., Li, X. & Porter, D. Twitter reveals human mobility dynamics during the covid-19 pandemic. *PloS one* **15**, e0241957 (2020).

34. Yilmazkuday, H. Stay-at-home works to fight against covid-19: international evidence from google mobility data. *J. Hum. Behav. Soc. Environ.* 1–11 (2020).

35. Pedregosa, F. *et al.* Scikit-learn: Machine learning in python. *J. machine Learn. research* **12**, 2825–2830 (2011).

36. Schnabel, R. B., Koonatz, J. E. & Weiss, B. E. A modular system of algorithms for unconstrained minimization. *ACM Transactions on Math. Softw. (TOMS)* **11**, 419–440 (1985).

37. Hellewell, J. *et al.* Feasibility of controlling covid-19 outbreaks by isolation of cases and contacts. *The Lancet Glob. Heal.* **8**, e488–e496 (2020).

38. Ferretti, L. *et al.* Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing. *Science* (2020).

39. Flaxman, S. *et al.* Estimating the effects of non-pharmaceutical interventions on covid-19 in europe. *Nature* **584**, 257–261 (2020).

40. Estrada, E. Covid-19 and sars-cov-2. modeling the present, looking at the future. *Phys. Reports* **869**, 1–51 (2020).

41. Maier, B. F. & Brockmann, D. Effective containment explains subexponential growth in recent confirmed covid-19 cases in china. *Science* **368**, 742–746 (2020).

42. Wong, G. N. *et al.* Modeling covid-19 dynamics in illinois under nonpharmaceutical interventions. *Phys. Rev. X* **10**, 041033 (2020).

43. Hernández-Orallo, E., Manzoni, P., Calafate, C. T. & Cano, J. Evaluating how smartphone contact tracing technology can reduce the spread of infectious diseases: The case of covid-19. *IEEE Access* **8**, 99083–99097 (2020).

44. Hernández-Orallo, E., Manzoni, P., Calafate, C. T. & Cano, J. Evaluating the effectiveness of covid-19 bluetooth-based smartphone contact tracing applications. *Appl. Sci.* **10**, 7113 (2020).

45. Khakharia, A. *et al.* Outbreak prediction of covid-19 for dense and populated countries using machine learning. *Annals Data Sci.* **8**, 1–19 (2021).

46. Hernandez-Matamoros, A., Fujita, H., Hayashi, T. & Perez-Meana, H. Forecasting of covid19 per regions using arima models and polynomial functions. *Appl. Soft Comput.* **96**, 106610 (2020).

47. Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S. & Ciccozzi, M. Application of the arima model on the covid-2019 epidemic dataset. *Data Brief* **29**, 105340 (2020).

48. Perone, G. An arima model to forecast the spread and the final size of covid-2019 epidemic in italy. *medRxiv* (2020).

49. Sahai, A. K., Rath, N., Sood, V. & Singh, M. P. Arima modelling & forecasting of covid-19 in top five affected countries. *Diabetes & metabolic syndrome* **14**, 1419–1427 (2020).

50. Lalmuanawma, S., Hussain, J. & Chhakchhuak, L. Applications of machine learning and artificial intelligence for covid-19 (sars-cov-2) pandemic: A review. *Chaos, Solitons & Fractals* **139**, 110059 (2020).

51. Rustam, F. *et al.* Covid-19 future forecasting using supervised machine learning models. *IEEE Access* **8**, 101489–101499 (2020).

52. Chimmula, V. K. R. & Zhang, L. Time series forecasting of covid-19 transmission in canada using lstm networks. *Chaos, Solitons & Fractals* **135**, 109864 (2020).

53. Ribeiro, M. H. D. M., da Silva, R. G., Mariani, V. C. & dos Santos Coelho, L. Short-term forecasting covid-19 cumulative confirmed cases: Perspectives for brazil. *Chaos, Solitons & Fractals* **135**, 109853 (2020).

54. Linka, K., Peirlinck, M. & Kuhl, E. The reproduction number of covid-19 and its correlation with public health interventions. *Comput. Mech.* **66**, 1035–1050 (2020).

55. Hernando, A., Mateo, D., Bayer, J. & Barrios, I. Radius of gyration as predictor of covid-19 deaths trend with three-weeks offset. *medRxiv* (2021).

56. Gonzalez, M. C., Hidalgo, C. A. & Barabasi, A.-L. Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008).

57. Cot, C., Cacciapaglia, G. & Sannino, F. Mining google and apple mobility data: temporal anatomy for covid-19 social distancing. *Sci. Reports* **11**, 4150 (2021).

58. Kraemer, M. U. G. *et al.* Data curation during a pandemic and lessons learned from covid-19. *Nat. Comput. Sci.* **1**, 9–10 (2021).

59. Castro, M., Ares, S., Cuesta, J. A. & Manrubia, S. The turning point and end of an expanding epidemic cannot be precisely forecast. *Proc. Natl. Acad. Sci.* **117**, 26190–26196 (2020).

Acknowledgements

This work has been partially supported by the Spanish Ministry of Science and Innovation, under grants RYC2018-025580-I, RTI2018-096384-B-I00, RTC-2017-6389-5 and RTC2019-007159-5, by the Fundación Séneca del Centro de Coordinación de la Investigación de la Región de Murcia under Project 20813/PI/18, by the "Conselleria de Educación, Investigación, Cultura y Deporte, Direcció General de Ciéncia i Investigació, Proyectos AICO/2020", Spain, under Grant AICO/2020/302 and a predoctoral contract by the Generalitat Valenciana and the European Social Fund under Grant ACIF/2018/219.

Author contributions statement

Conceptualization, S.G.C. and J.L.E; methodology, S.G.C. and J.L.E.; software, J.M.G., R.M.E., A.B.C, E.H.O.; validation, S.G.C., J.L.E., R.M.E. and J.M.C.; formal analysis, S.G.C., R.H.S., R.M.E., J.L.E., A.B.C,; investigation, S.G.C., R.H.S., J.L.E., R.M.E., A.B.C., E.H.O. ; resources, S.G.C. and J.M.G. ; data curation, S.G.C., J.M.G., R.H.S. and R.M.E.; writing—original draft preparation, S.G.C., J.M.C.; writing—review and editing, J.M.G., E.H.O.; visualization, J.M.G., R.H.S., A.B.C, ; supervision, J.L.E and J.M.C.; funding acquisition, J.M.C. All authors have read and agreed to the published version of the manuscript.

Figures

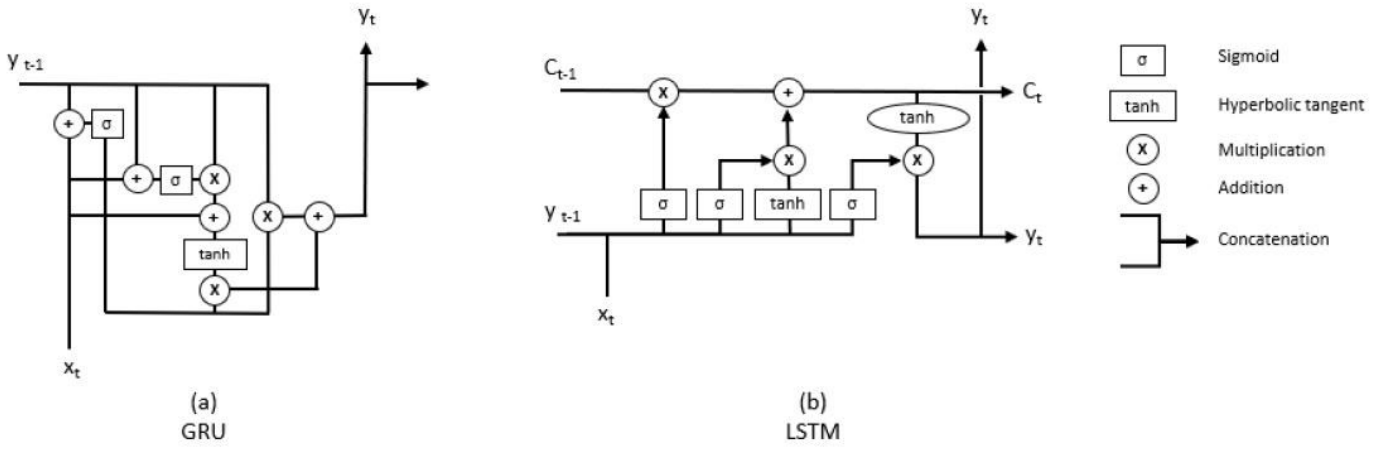


Figure 1

Diagram of a GRU and LSTM unit. Where xt represents the input and yt the forecast in a step (yt-1 for forecast in the previous steps). For LSTM, the Ct indicates the state that is passing from one LSTM unit to another.

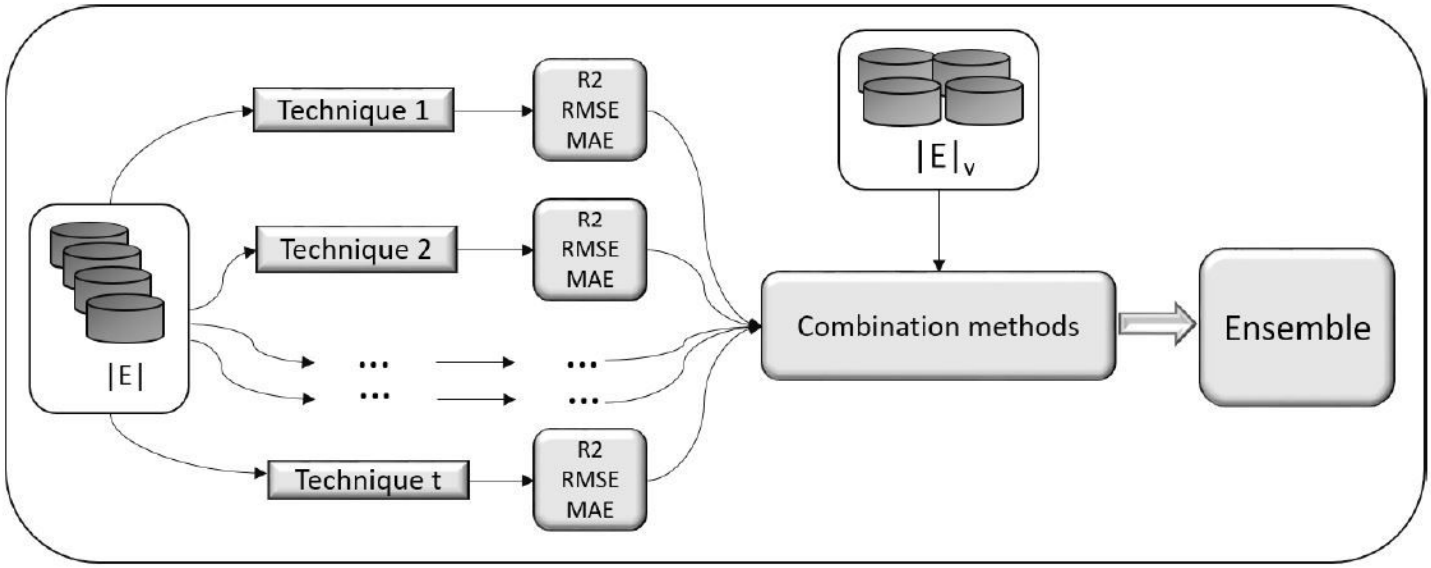


Figure 2

Outline of the proposed ensemble approach.

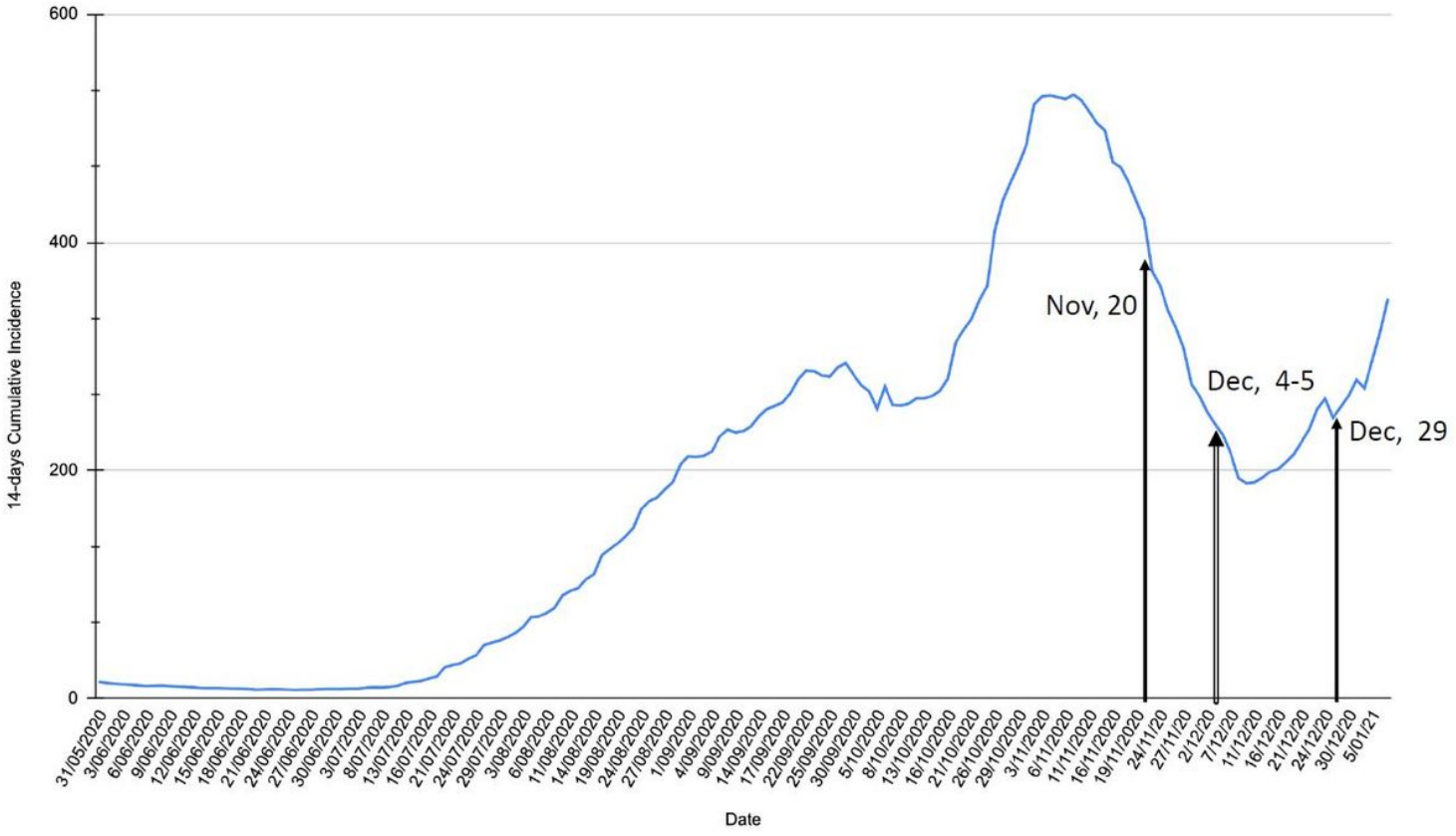
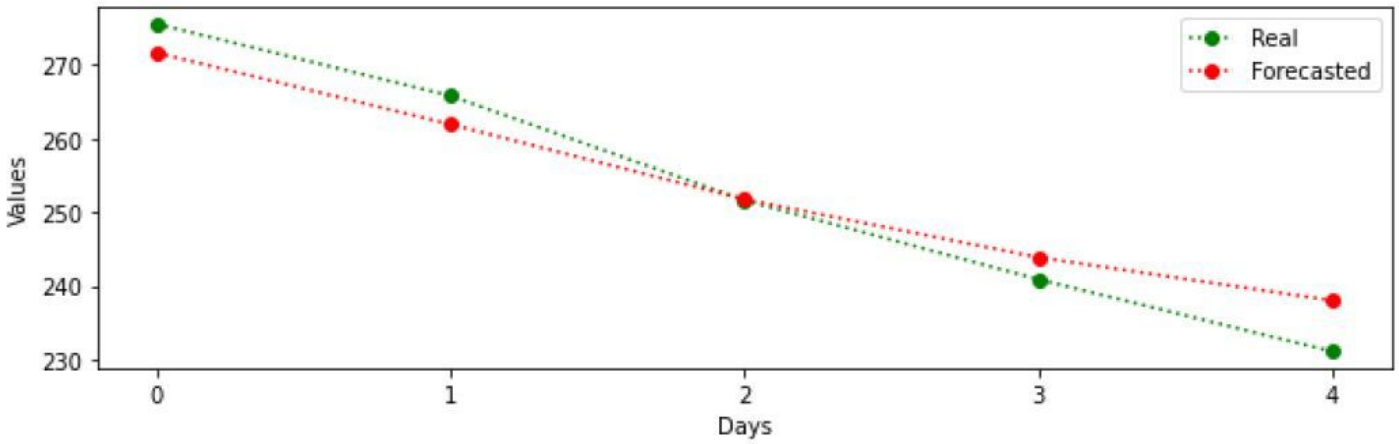
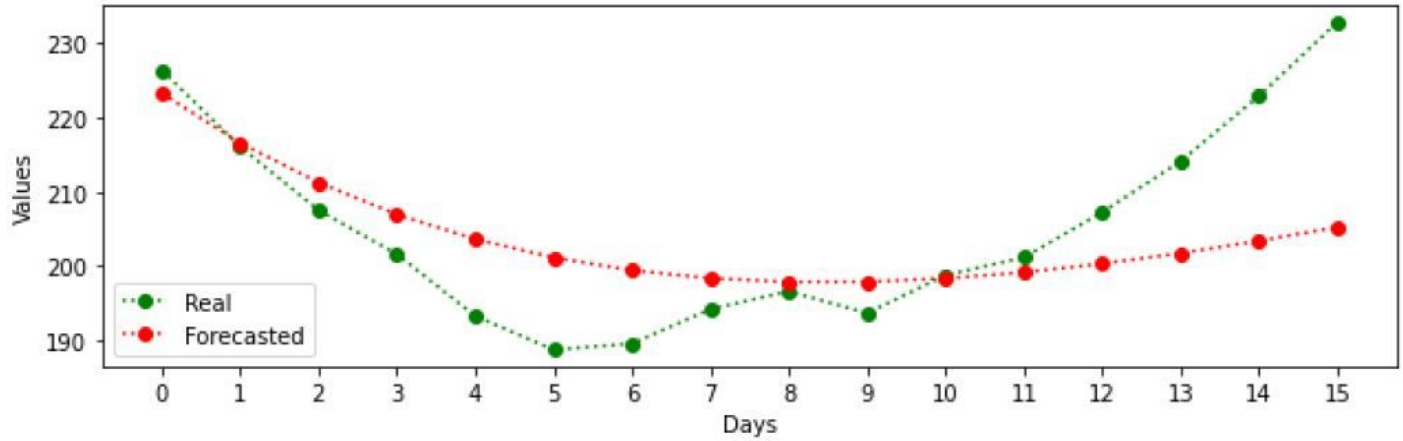


Figure 3

14-day cumulative incidence (CI) in Spain. The evaluation dates are highlighted to let the reader know the trend of 14-day CI at that period.



(a) DATASET 1



(b) DATASET 2

Figure 4

14-day CI accuracy prediction for both datasets.

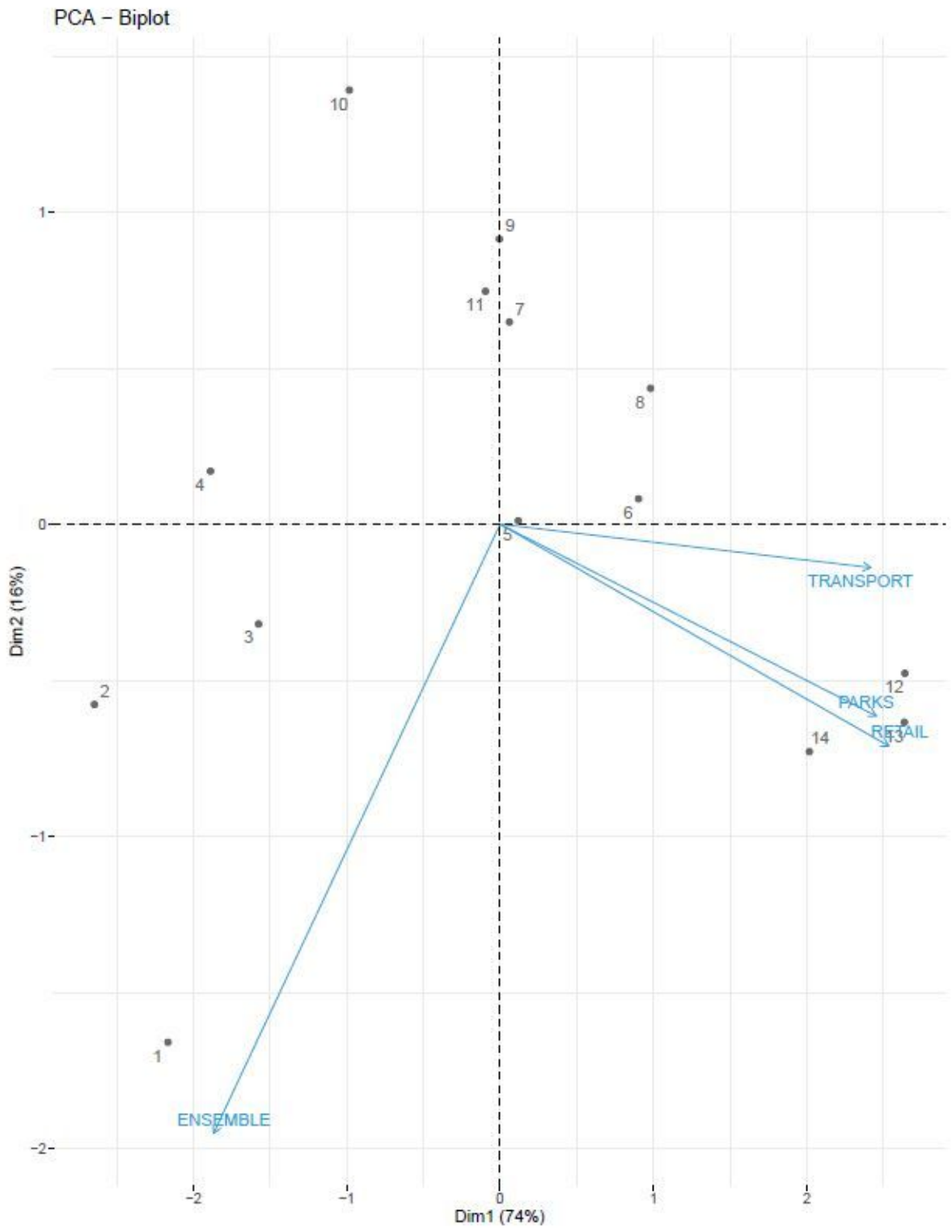


Figure 5

PCA to ensemble approach and mobility variables. Positively correlated variables point to the same side of the plot. Negatively correlated variables point to opposite sides of the graph.

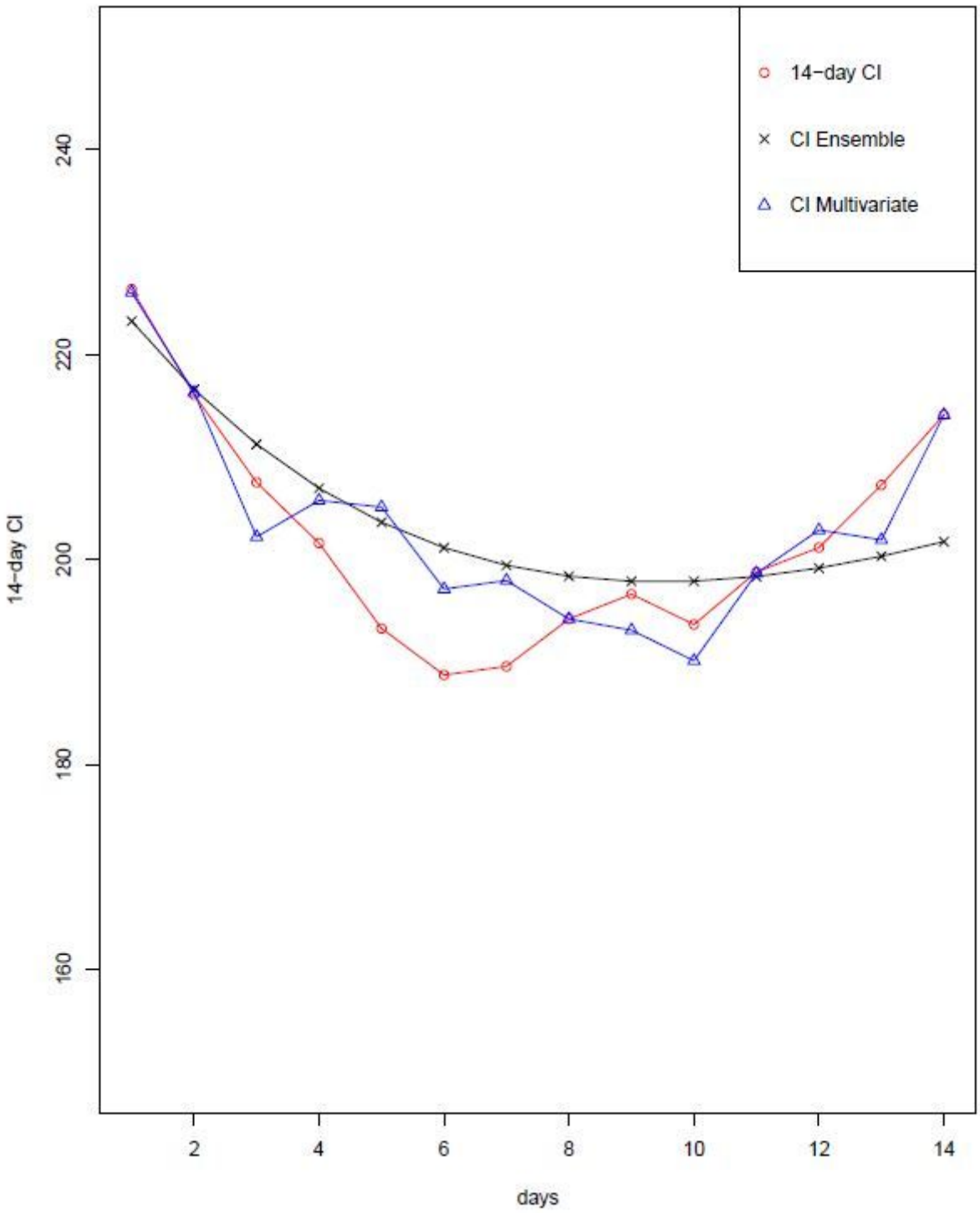


Figure 6

14-day CI accuracy prediction for different estimated models.