Forecasting COVID-19 daily cases using phone call data

Bahman Rostami-Tabar [a],*, Juan F. Rendon-Sanchez [b],*

[a] Cardiff Business School, 3 Colum Drive, CF10 3EU, Cardiff, UK
[b] Cardiff School of Computer Science and Informatics, Queen's Buildings, 5 The Parade, Roath, CF24 3AA, Cardiff, UK

ARTICLE INFO

ABSTRACT

The need to forecast COVID-19 related variables continues to be pressing as the epidemic unfolds. Different efforts have been made, with compartmental models in epidemiology and statistical models such as AutoRegressive Integrated Moving Average (ARIMA), Exponential Smoothing (ETS) or computing intelligence models. These efforts have proved useful in some instances by allowing decision makers to distinguish different scenarios during the emergency, but their accuracy has been disappointing, forecasts ignore uncertainties and less attention is given to local areas. In this study, we propose a simple Multiple Linear Regression model, optimised to use phone call data to forecast the number of daily confirmed cases. Moreover, we produce a probabilistic forecast that allows decision makers to better deal with risk. Our proposed approach outperforms ARIMA, ETS, Seasonal Naive, Prophet and a regression model without call data, evaluated by three point forecast error metrics, one prediction interval and two probabilistic forecast accuracy measures. The simplicity, interpretability and reliability of the model, obtained in a careful forecasting exercise, is a meaningful contribution to decision makers at local level who acutely need to organise resources in already strained health services. We hope that this model would serve as a building block of other forecasting efforts that on the one hand would help front-line personal and decision makers at local level, and on the other would facilitate the communication with other modelling efforts being made at the national level to improve the way we tackle this pandemic and other similar future challenges.

1. Introduction

Since its discovery at the end of 2019 in Wuhan, China, the spread of coronavirus (SARS-CoV-2) has shaken one country after another putting governments, infrastructure, local and international cooperation to the test as it has claimed the lives of more than a million people worldwide [1] and caused severe disruptions to every-day life and the economy.

The epidemic is evolving with regional and local differences. The way it spreads, as pointed out by Hamzah et al. [2], is largely influenced by each country's policies. Consequently, in the UK, some of the regional differences might be strongly linked to differences in approaches or policies or their timing to tackle the epidemic in England, Wales, Scotland and Northern Ireland. Additionally, other factors, linked to human direct interaction and exposure, such as life conditions (including income, accommodation, vulnerability), engagement in economic activities with close interaction and mobility patterns can further create differences in the evolution of the epidemic at regional and local levels. These complexities are being researched in interdisciplinary areas such as digital epidemiology [3,4].

As the epidemic unfolds, decision makers face the difficulties posed by uneven developments and model limitations to predict its evolution. Increasing numbers of confirmed cases lead to an increase in demand for hospital beds, and in the most severe cases also ICU beds. The uneven spread of the infection, manifested in large variations at local level, render important to monitor its spread and attempt to predict the number of new cases. The ability to accurately forecast the number of positive confirmed cases few weeks in advance, would allow local decision makers to put in place all measures required to increase capacity in hospital beds to treat COVID-19 patients before it occurs, instead of reacting to it.

This study has initially been motivated by a real-life forecasting problem faced by a local country council in the National Health Service (NHS) in England. The problem highlights the lack of a rigorous method to forecast the number of confirmed cases at ground/local level. Decision makers at the local level need forecasts to: (i) inform general strategy; (ii) adjust hospital bed capacity; (iii) cancel/postpone non-urgent hospital operations; (iv) redirect available medical resources to COVID-19 wards; and (v) plan for extra capacity in body storage of mortuary services.

Additionally, some limitations in the literature have also motivated the development of our study which are summarised as follows: (i) the forecast accuracy of models is not systematically reported; (ii) there are big differences between predictions and what happened in reality and forecast errors greatly vary; (iii) in some studies it is unclear how forecasts are generated and evaluated, e.g. whether a rolling window is used and what is the forecast horizon; and (iv) most of the studies focus on point forecasts and ignore the uncertainty associated with them, which may lead to less effective decisions. Despite the limitations of these efforts, the need to have appropriate tools to aid decision making continues to be pressing and sensible contributions within forecasting are sought after, as evidenced by the publication, already since early stages of the outbreak [5]. Frustration with regards to forecast accuracy when forecasting COVID-19 has also been summarised by Ioannidis et al. [6]. A myriad of reasons has been given, including poor data, lack of incorporation of epidemiological features and lack of transparency. It is important to note that the communication between decision makers at local and national levels is also acutely needed. Therefore, any experience with different forecasting and other decision tools at local level with a clear assessment of their accuracy will facilitate the integration of models and communication as the epidemic continues to unfold.

This paper contributes to the literature by developing a probabilistic forecasting approach that uses phone call data to forecast COVID-19 cases at a local level. Overall, the forecasting approach, accuracy assessment, and empirical conditions in our present work are additions to the increasing body of knowledge about foresting this pandemic, and its products (models, assessment and conditions) can be used as building blocks in broader approaches and frameworks that encompass diverse models. It is a contribution towards enlarging the set of tools and accuracy assessment, to aid front line health personnel and managers facing the crisis in taking decisions on resources and operations.

Our contributions are fourfold: (i) we propose a novel forecasting model to accurately forecast daily COVID-19 confirmed cases; (ii) we examine the use of phone call data and provide evidence of their usefulness in forecasting daily confirmed cases; (iii) we provide probabilistic forecasts that quantify uncertainties in future confirmed cases; (iv) we benchmark the accuracy of our model against five techniques including AutoRegressive Integrated Moving Average, ARIMA [7], Exponential Smoothing State Space, ETS [8], Seasonal Naive, Snaive [9], Prophet [10] and a regression model without considering call data.

The rest of the paper is organised as follows: Section 2 provides a brief overview of the efforts to forecast COVID-19, Section 3 starts by analysing data to then introduce the models used in this study and the forecast performance evaluation scheme and metrics. Section 4 presents the results of the study, followed by concluding remarks in Section 5.

2. Research background: forecasting COVID-19

Forecasting COVID-19 has attracted a lot of attention of researchers and many studies have analysed various related variables. Some studies focus on the global and national level; others focus on regional level and few consider the local level. The following is an overview of approaches used to forecast different variables of COVID-19 to date.

Petropoulos and Makridakis [11] forecast the cumulative number of daily confirmed cases globally with exponential smoothing models. Since this early effort, the limitations of the models were clear and acknowledged by the authors and attributed to strong government interventions and accuracy of data. Benvenuto et al. [12] predict the epidemiological trend of incidence and prevalence of COVID-19 worldwide using ARIMA models. Feroze [13]

use Bayesian Structural Time Series Models [14] to forecast the daily total number of positive cases in USA, Brazil, Russia, India and UK for the next 30 days. Authors report a better performance from their models when compared to ARIMA models.

At country level, Fong et al. [15] proposed a methodology for fitting models with scarce data at the beginning of the pandemic with data from China. A polynomial neural network with corrective feedback is found to perform better, according to RMSE (Root Mean Square Error), in a panel selection framework including other polynomial neural networks, ARIMA, Linear Regression, Holt-Winters models, Support Vector Machines and Fast Decision Tree, among others. In a similar approach, Fong et al. [16] use Monte Carlo simulation to forecast the total daily cost associated to the pandemic by using scarce data. Two groups of variables, one with time-series data and other with scalar estimates, are used as inputs, with several models being subject to a panel selection process. In the last stages, visualisation through fuzzy rules is proposed to enhance decision-making. This is an attempt to tackle data scarcity and uncertainties, with the aim of supporting decision making in an exercise of broad scope. The approach, tested with data COVID-19 data from China, provides, according to the authors, better qualitative results for decision making than individual forecasting techniques alone. The approaches in these two articles are first steps towards forecasting frameworks to inform decision making during pandemics. An interesting next step forward would be to use open source forecasting tools, to equate the use of open data, in order to promote the construction of such frameworks.

Yousaf et al. [17] forecast daily confirmed cases with ARIMA models at country level in Pakistan. Of the same nature is the research by Moftakhar et al. [18] with data from Iran, in which observed new cases are used to predict the number of patients. Artificial Neural Networks (ANN) and ARIMA are used, with the later showing better accuracy. ARIMA models are also used by Gupta and Pal [19] to forecast infection cases in India for the next 30 days. Harun et al. [20] forecast the number of COVID-19 cases in Germany, United Kingdom, France, Italy, Russia, Canada, Japan, and Turkey by using linear and non-linear regression models, ARIMA and exponential smoothing methods. Different best performing forecasting models are identified for different countries.

Hazarika and Gupta [21] and references therein show that the proposal of modified machine learning models has been very active. The authors propose a hybrid Random Vector Functional link (RVFL) network and a wavelet-coupled RVFL (WCRVFL) network to forecast the number of infected people in Brazil, India, Peru, Russia and the USA. Performance is evaluated when forecasting 60 days ahead, against support vector regression (SVR) models and the conventional RVFL.

Tomar and Gupta [22] use a long short-term memory (LSTM) recurrent neural network to forecast daily and total positive cases, total recovered and total deceased in India. A simple curb fitting with an exponential model is used to assess the effect of a lock-down and social distancing. LSTM networks were also used with data from Canada by Chimmula and Zhang [23].

Perc et al. [3] forecast the number of daily confirmed cases in United States, Slovenia, Iran and Germany with an iterative model that uses the average growth of cases and takes into account the recovered and deceased. By devising the model around growth and targeting its use to the short term, it was rendered helpful even in the absence of a systematic accuracy study.

Logistic models of the SEIRD type, which divide population in groups (Susceptible, Exposed, Infected, Removed and Dead), and require the solution of a system of differential equations, have been used abundantly, with different variants, to describe and predict the evolution of the epidemic globally [2] and in regions and countries [24–27]. Martelloni and Martelloni [24] apply a

SIRD model to Italy and adapt it to account for the introduction of lock-downs by governments, the influence of asymptomatic and the absence of a criteria to define the susceptible group. Fanelli and Piazza [25] use a SIRD model to explore the temporal dynamics of the epidemic in China, Italy and France. Sarkar et al. [28] adapt the SIRD model to account for asymptomatic, isolated infected and quarantined susceptible. Mohd and Sulayman [26] propose a modified logistic model to account for reinfections, false detection problems and scarcity of medical equipment in the context of a developing country and found that in such circumstances there can be unstable scenarios where decisions based on the reproduction number R_0 should be made with caution. In some studies [24,25] there are hints of universal characteristics, whereas in Mohd and Sulayman [26] the particular circumstances of a developing country suggests that behaviours of the epidemic might have important country differences.

Sardar et al. [29] investigate several models covering the mathematical (SIRD) and statistical types, to study the effect of lock-down in India: a SIRD-type logistic model, expanded further than Sarkar et al. [28] that accommodates for lock-down; an ARIMA model; an exponential smoothing model with ARMA errors, trend and seasonal components (TBATS); a hybrid statistical model based on the combination of ARIMA and TBATS and a weighted combination of the SIRD model and the best statistical one. Daily notified cases from five states in India and the overall country are used to fit the models, which helped assessing different scenarios determined by the effective reproduction rate number.

Sophisticated spatio-temporal models have been explored to project infections, deaths and ICU beds and hospital admissions needed in the US [30–32]. This and related work [33] are signalling the need of complex methods for some settings. A trend that is also manifested, for example, in the investigation of ARIMA models in conjunction with spacial relations between neighbouring countries [34].

The survey by Shinde et al. [5], focuses mostly on research at the country level which considers two categories for data sources: i) official government, and ii) trusted organisations, such as WHO, and social media. The survey also classifies models into two groups: i) mathematical/analytical, and ii) machine learning/data science. These classifications continue to be meaningful to-date. It is interesting to notice the way in which phone data has been use. Mobile phone data has been used to collect statistics of movement of people and data from search engines and newspaper reports has been used to predict the total number of deceased. The scarcity of research on using data from phone calls to health services persists to-date.

At regional level, Anastassopoulou et al. [35] propose a methodology to estimate epidemiological parameters and apply a SIRD model to forecast the number of infections, recovered and deceased in the region of Hubei, China. After a coarse estimation of parameters, a non-linear optimisation routine is used to obtain refined estimates. The official figures from the region fell inside the upper and lower bounds, albeit very wide, that had been produced by authors. Despite the inaccuracy, the lessons learned in how to calibrate this type of models are very valuable.

Massonnaud et al. [36] adapt a SEIR model to formulate scenarios for French metropolitan regions based on different reproduction rates. The daily number of COVID-19 cases, hospitalisations and deaths, the needs in ICU beds per region and the reaching date of ICU capacity limits were estimated. Hospital catchment areas are used and then aggregated by French region.

Weissman et al. [37] implement a SIR model to plan scenarios (defined by doubling times of 2, 6 and 10 days) in 3 hospitals in greater Philadelphia region, US. Three variables are predicted: hospital capacity, patients requiring ICU beds and patients requiring ventilators. The accuracy of the predictions is not reported,

but the usefulness of the scenarios in the planning ahead is recognised. This assessment and the construction of models is facilitated by a close collaboration between operational, medical and data personnel.

When considering this overview of forecasting efforts, it is clear that accuracy has been disappointing in many cases. The majority of the studies focus only on point forecasts and ignore the uncertainty that may lead to less effective decisions. In some cases, the forecast accuracy has not been reported or it is unclear how it has been evaluated. It is evident the effort in proposing and testing machine learning, statistical and mathematical models. In the area of machine learning there is a emphasis on hybridisation to cope with complexity whereas in the area of SIRD models there is an attempt to adapt them to different circumstances and government interventions. However, limitations of different models might not be reduced unless connected with other approaches such as spatio-temporal complex models like Global Epidemic and Mobility Model (GLEAM) [32], and well articulated to policy making. It is also evident that phone data from official help lines has not been extensively explored, or at least reported.

3. Experimental design

3.1. Data

Data used in this paper comprised the number of daily COVID-19 confirmed cases and the number of daily phone calls received at the National Health Service 111 (NHS 111) at one of the largest Non-metropolitan country council in the East Midlands region of England between 18 March 2020 and 19 October 2020. Data is publicly available and extracted from the Public Health UK,[1] and NHS Digital[2] respectively.

Fig. 1 illustrates the time plot of the confirmed cases. Although the time series is noisy, some systematic patterns are visible. The time plot shows a trend and the number of positive cases in the weekend are lower than the working days. We also analyse the Autocorrelation (ACF) and Partial Autocorrelation (PACF) Functions to investigate the link between the number of confirmed cases and its lagged values. Fig. 2 shows the ACF (Fig. 2-A) and PACF (Fig. 2-B) of the time series. The ACF plot highlights a trend because by increasing the lag, the value of ACF decreases exponentially. Moreover, PACF reveals that there are some positive lags that need to be considered in the forecasting model such as $lag\,1$, $lag\,2$, $lag\,8$, and $lag\,10$.

Fig. 3 presents the time plot of the NHS 111 calls highlighting the presence of a trend. The data is less noisy and there is no seasonal pattern. NHS 111 call data can be used as a predictor of the number of daily confirmed cases, therefore, we need to produce its forecast.

In the relationship between the number of confirmed cases and the number of NHS 111 calls on day t, the former may be related to past lags of the later. The sample Cross-Correlation Function (CCF) can be used to identify lags of the number of NHS 111 calls series that might be useful predictors of the number of confirmed cases. For instance, consider $lag = -24$, the CCF value will give the correlation between the number of NHS 111 calls on day $t - 24$ and the number of confirmed cases on day t.

Fig. 4 presents the linear relationship between confirmed cases and NHS 111 lags. We are interested in the negative lags, because we use NHS111 calls to predict the number of confirmed cases. It seems that confirmed cases are relatively high almost 0–24 days after high NHS 111 calls (i.e., significant correlation at lags of 0 to 24). There are a lot of models that we could try based on the CCF,
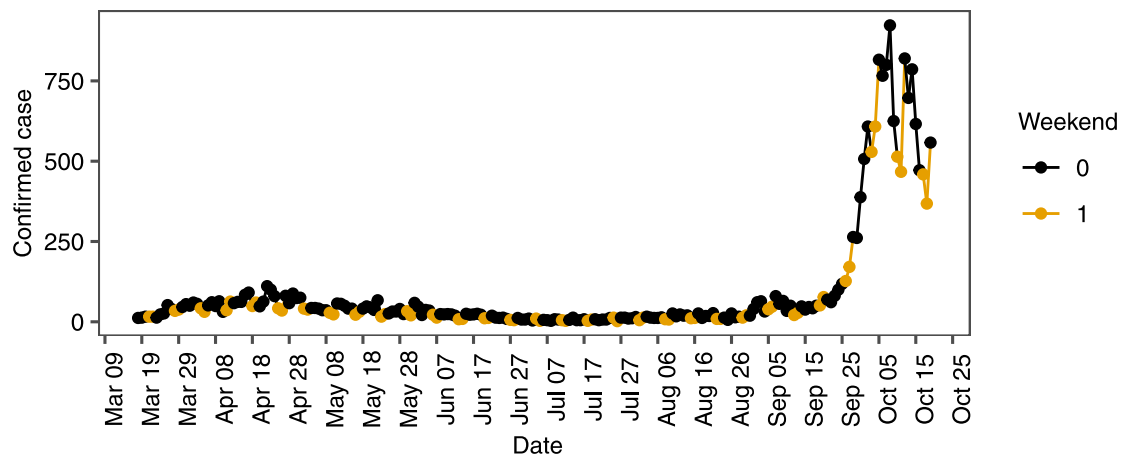
---

[1] https://coronavirus.data.gov.uk/.
[2] https://digital.nhs.uk/coronavirus.

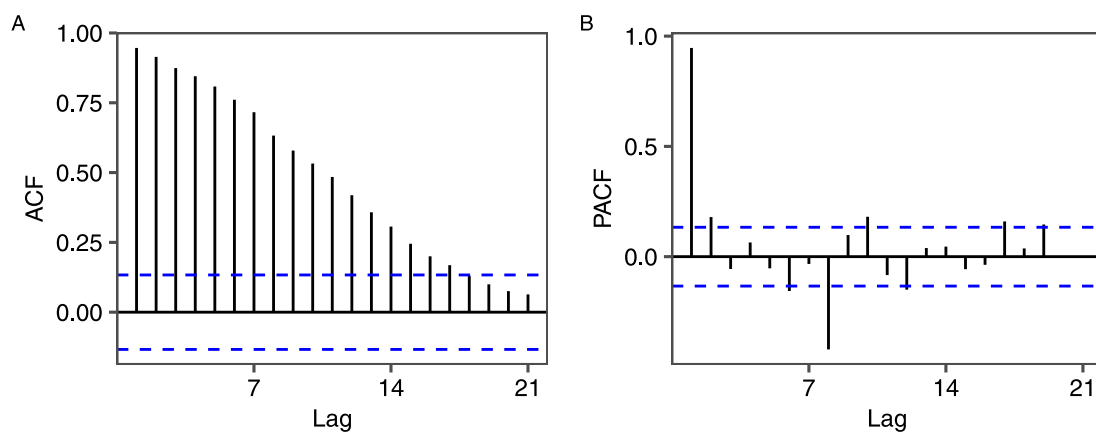**Fig. 1.** Time series of daily confirmed cases.

Fig. 2. Autocorrelation and partial autocorrelation of the time series of confirmed cases.
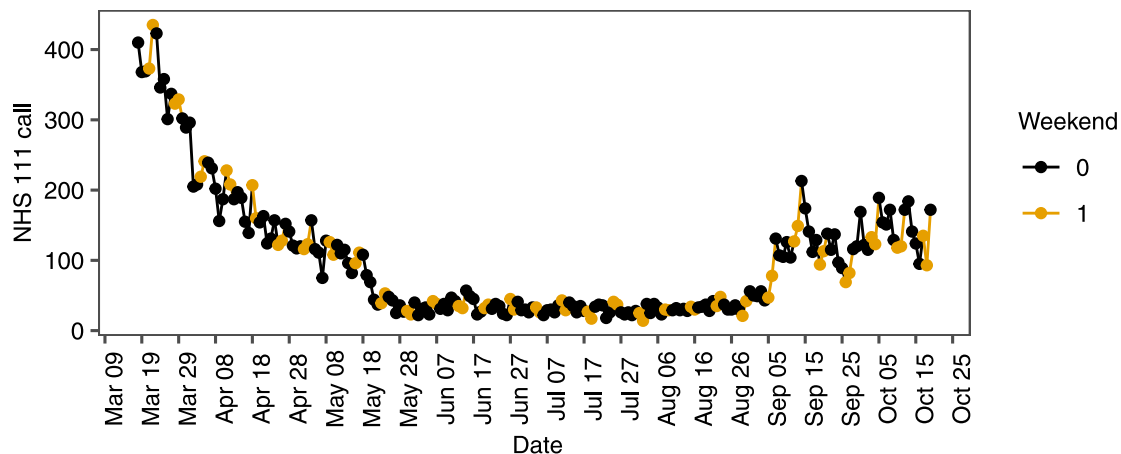

Fig. 3. Time series of daily phone calls.

ACF, PACF and the presence of trend and the effect of weekends for these data. In the next section, we propose a Multiple Linear Regression model in which the number of confirmed cases, is a linear combination of these features.

3.2. Proposed model, multiple linear regression with call data (MLR_T)

In this section, we model the number of confirmed cases as potentially a function of past lags of confirmed cases, current and past lags of the number of NHS 111 calls, the weekend and the trend. These components allow for possible interpretations as described in Section 3.1.

We denote Y_t as the number of positive confirmed cases on day t. The predictors are presented using the following notations:

- $NHS111_t$: number of NHS 111 calls on day t.
- $trend_t$: local trend on day t.
- Y_{t-k}: number of positive confirmed cases on day $t - k$.
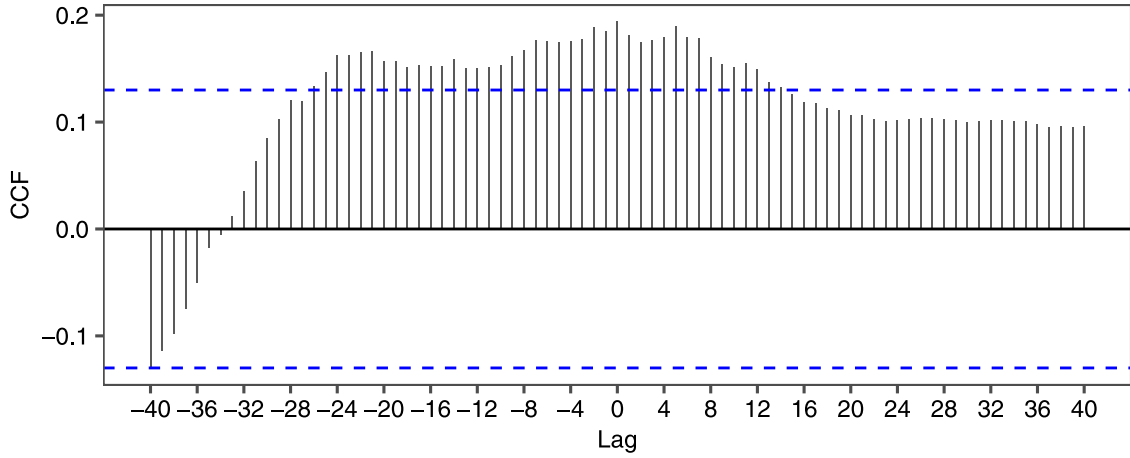- $NHS111_{t-k-4}$: NHS 111 calls on day $t - k - 4$.

Fig. 4. Cross-correlation of the past values of NHS 111 calls and the number of confirmed cases.

- $Weekend_t$: Weekend dummy; $Weekend_t$ is 1 if t falls on a Weekend and 0 otherwise.

$$Y_t = \beta_0 + \beta_1 \text{NHS } 111_t + \text{trend}_t + \underbrace{\sum_{k=1}^{21} \beta_{k+1} Y_{t-k}}_{\text{AutoRegressive effect}}$$

$$+ \underbrace{\sum_{k=1}^{26} \beta_{22+k} \text{NHS } 111_{t-k-4}}_{\text{NHS 111 lag effect}} + Weekend_t + \varepsilon_t \qquad (1)$$

The mathematical representation of the proposed model is shown in Eq. (1). We should note that, to ensure achieving strictly positive values for the number of confirmed cases, we have used the `log()` transformation for the response variable, confirmed cases, and its lagged values as predictors. Specified transformation is automatically back-transformed to produce forecasts.

Not all potential predictors are included in the final forecasting model. After building the model, we need to select the predictors that are useful in producing accurate forecasts. To that end, we have conducted a stepwise regression and select predictors that improve the Out-of-Sample forecast accuracy based on RMSE and Continuous Ranked Probability Score (CRPS). Due to the amount of uncertainty in the number of positive confirmed cases, we produce the estimated probability distribution of the number of positive confirmed cases in future time periods in addition to the point forecast. This allows us to provide more information about the forecasts that could be fruitfully used by decision makers.

The proposed model uses the NHS 111 call, its lags and the lagged values of confirmed cases as predictors, therefore, in order to produce ex-ante forecasts [9], we need to generate forecasts of these predictors. Fig. 5 depicts the steps of the forecasting process, implemented in R.

To forecast NHS 111 calls, we examined ARIMA, ETS and their combination. We use forecasts generated by ARIMA for NHS 111 calls as it results in more forecast accuracy. For the lagged values of confirmed cases, we examines Naive [9], ARIMA, ETS and their combinations. Forecasts generated by ETS result in more accurate forecasts overall, however, forecasts from `Naive` method were accurate for shorter horizons. Therefore, we replace the forecasts generated from ETS by those generated from the `Naive` method when $h < 5$. We have compared various horizons, h and this yields in a more accurate forecast of confirmed cases in the Out-of-sample.

Fig. 6 illustrates an example of the probabilistic forecast generated by the proposed model.

3.3. Benchmarks

We use five different benchmark methods to compare their forecast accuracy with our proposed model. We consider Exponential smoothing and ARIMA models as the two most widely used approaches to time series forecasting, in addition to a regression approach without using call data, Seasonal Naive and Prophet. These benchmark methods are used to show whether there is any adding value in using the proposed model or not.

3.3.1. Proposed model without call data (MLR_W)

In order to show the impact of incorporating the phone call data in the modelling and its effect on forecast accuracy improvement, we use the same model proposed in Section 3.2 without including NHS 111 calls and its lagged variables.

3.3.2. Exponential smoothing state space model (ETS)

The second benchmark is the automatic exponential smoothing [9] model that is incorporated using corresponding implementation of the fable package in R. We use the ETS() function in the fable package [38] to generate daily forecast of the confirmed cases.

3.3.3. Autoregressive integrated moving average (ARIMA)

The third benchmark model is an automatic ARIMA model [9]. The orders of the model and parameters are selected by minimising the corrected AIC (Akaike information criterion). We use the implementation of automatic ARIMA in `fable` package; ARIMA() function is used to generate the forecasts of daily confirmed cases [38].

3.3.4. Prophet

Prophet is a forecasting procedure created by Facebook [10] that accounts for multiple seasonality and piecewise trend. Prophet works well on daily data, is robust to missing data and shifts in the trend, and typically handles outliers well. It is popular and automated which makes it easy to learn and use in practice. The model is incorporated using corresponding implementation of the Fable package in R. We use the prophet() function in the fable package to generate daily forecast of confirmed cases [38].

3.3.5. Seasonal Naive

The Seasonal naive model is a simple method in which forecast of confirmed cases is equal to the last observed confirmed cases from the same day of the previous week. This method is useful for time series with seasonality.
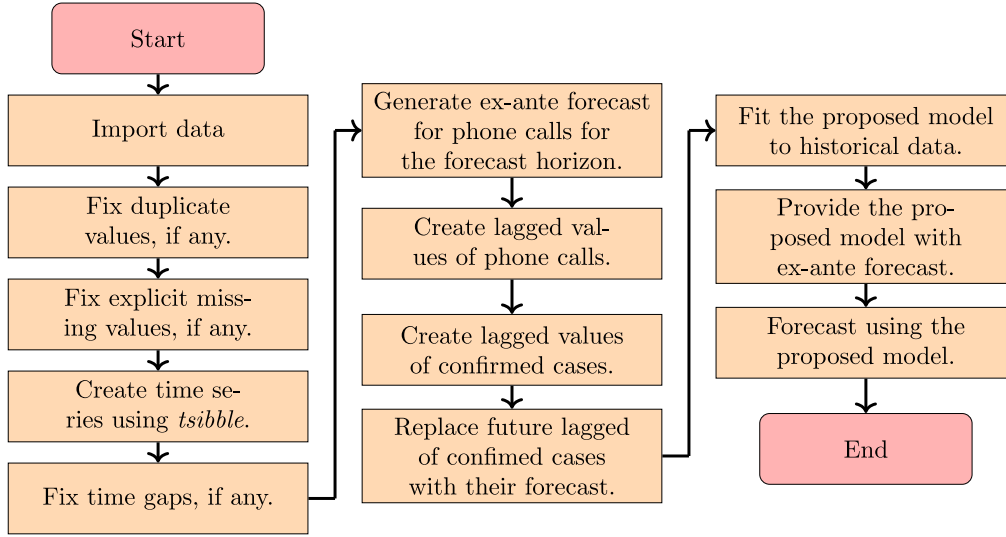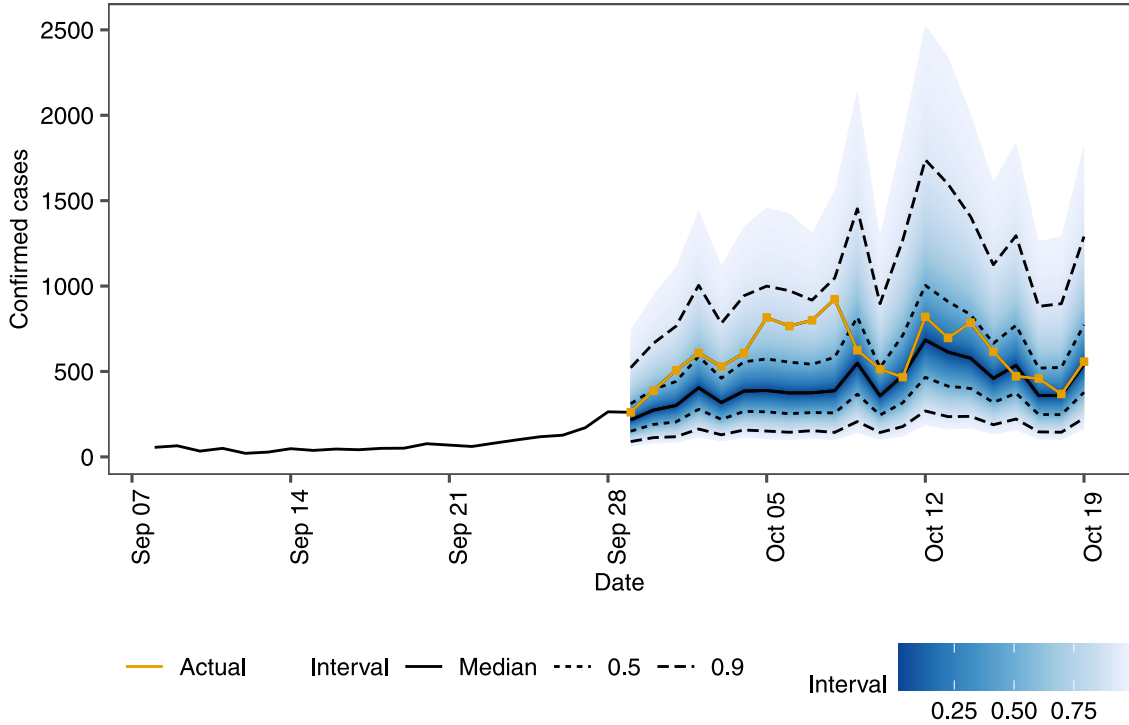
Fig. 5. Steps of the forecasting process.

Fig. 6. Sample forecast produced by the proposed model.

3.4. Forecast performance evaluation schemes and metrics

We first split the dataset into training (70%) and test (30%) sets. We apply the forecasting models on the training set and evaluate the forecast performance of the proposed MLR approach and benchmarks on the test set. Evaluation is conducted using a rolling origin forecasting study with re-estimation. Forecasting horizon is chosen to be $h = 21$ which corresponds to three weeks. This allows decision makers to use forecast to inform decisions on resource allocation and planning in the health service. We have considered six different forecast accuracy metrics: i) three point forecast error metrics, ii) one prediction interval accuracy measure, and (iii) two probabilistic forecast accuracy metrics.

3.4.1. Point forecast accuracy measures

The point forecasting metrics are the ME (Mean Error), MAE (Mean Absolute Error) and RMSE (Root Mean Square Error) for point forecasts [39]. These measures help us to evaluate the performance of models from different perspectives. ME provides the overall direction of the error. It reveals whether the produced forecasts on average are too high or too low. MAE is strictly appropriate for the median, i.e. if the MAE of a model is smaller than others, it produces forecasts closer to the median of the data than the others. Finally, RMSE is minimised for the optimal mean forecast. That means a model with the lower value of RMSE produces more accurate mean values comparing to others. We evaluate the ME, MAE and RMSE for each model and each forecasting step, separately. We also report the overall performance

of each model across all horizons. A model with an error measure closer to zero is better.

3.4.2. Prediction interval accuracy measures

An appropriate prediction interval accuracy measure should account for both the coverage and the width of the prediction interval. Winkler [40] proposed a measure to calculate the accuracy of the forecast prediction interval for any given model.

If y_t is the observation at time t and $[l_t, u_t]$ represents the $100(1 - \alpha)\%$ prediction interval at time t, then the score for each period is:

$$W(l_t, u_t, y_t) = \begin{cases} u_t - l_t & \text{if } l_t < y_t < u_t \\ (u_t - l_t) + \frac{2}{\alpha}(l_t - y_t) & \text{if } y_t < l_t \\ (u_t - l_t) + \frac{2}{\alpha}(y_t - u_t) & \text{if } y_t > u_t \end{cases} \tag{2}$$

The Winkler score is the average of scores across all periods. We prefer a model that has a narrow prediction interval with high coverage, therefore a model with smaller score is better. In this study, Winkler score is calculated based on 95% prediction intervals.

3.4.3. Probabilistic accuracy measures

In addition to the point and prediction interval accuracy measures, we also report two probabilistic forecasting measures: (i) Percentile score and , (ii) Continuous Ranked Probability Score (CRPS).

The percentile score is a strictly appropriate evaluation criterion for quantiles. The evaluation is performed on a dense probability grid for all percentiles $(1\%, \ldots, 99\%)$ [41]. For each time period t, we obtain the quantile, $q_{i,t}$, where $i = 1, 2, \ldots, 99$. Then, the percentile score is given by the pinball loss function at each period t:

$$L(q_{i,t}, y_t) = \begin{cases} (1 - i/100)(q_{i,t} - y_t) & y_t < q_{i,t} \\ (i/100)(y_t - q_{i,t}) & y_t \geq q_{i,t}. \end{cases} \tag{3}$$

The score can then be calculated for each forecast horizon and across all percentiles. We also report the overall percentile score across all horizons. If the observations does not deviate from the forecast distribution, then the average score is small. A model with smaller percentile score is better.

CRPS summaries the quality of a continuous probability forecast with a single score by measuring the distance between the forecast and the observed cumulative Distribution Function. It reveals how closely the CDF of the forecast matches that of the corresponding observations.

We should note that in order to report the result of the performance evaluation using time series cross validation, we have averaged the score of each measure computed for different rolling origins.

4. Result and discussion

In this section, we present the forecast accuracy of our proposed MLR model and its benchmarks for point forecast, prediction interval and probabilistic forecasts using out-of-sample data using time series cross validation approach.

Table 1 presents the overall performance of forecasting models for three types of forecast accuracy measures: (i) point error measures using ME, RMSE, MAE; (ii) prediction interval accuracy measure using Winkler score, and (iii) probabilistic accuracy measures using Percentile and CRPS. The reported scores are averaged across all rolling origins and forecast horizons. Overall, the proposed MLR model clearly outperforms benchmarks for all forecast accuracy measures.

We first observe the performance of forecasting models using the point forecast accuracy measures (columns 2–4). ME shows

the forecast bias of considered models which is the tenancy of a forecast to be consistently higher or lower than actual values. Given that ME is positive for all models, generated forecasts are positively biased manning that all forecasting models are overestimating the daily number of confirmed cases on average. The MLR_T model has the lowest bias comparing to all benchmarks. Decision makers should consider this when making decisions based on the forecasts.

$RMSE$ is the squared root of MSE which is minimised by mean. If the decision maker is interested in producing the expected average of the daily confirmed cases, and then identify which model outperforms benchmarks in terms of the expected values, then RMSE measure is an appropriate indicator. Table 1 indicates that MLR_T outperforms benchmarks based on $RMSE$.

MAE accuracy measure is another indicator that is useful to evaluate the point forecast accuracy which is minimised by median. That means, if the decision maker is interested in identifying which forecasting model generates more accurate forecast of daily confirmed cases in terms of median values, then MAE should be used. Results indicate that MLR_T model outperforms others based on MAE.

In addition to point forecast accuracy, we believe that the forecast should acknowledge the uncertainty of the future confirmed cases and it should be an integral part of the decision-making process. The report of uncertainty allows for a more informed decision and better risk management strategy. We first report forecast uncertainty using a Winkler (column 5). Results shows that the proposed model has the lowest Winkler score, which means on average it has more accurate coverage and narrower intervals than all benchmarks.

Most importantly, we provide the accuracy of the probability forecast distribution that contains all forecast properties including mean, median, prediction intervals and quantiles. This evaluates how well the forecast distribution is capturing the structure of the data. Results indicate that the MLR_T model provides more accurate probability forecast distribution than others. It means that the proposed model can inform decisions more accurately about uncertainty than others.

In addition to the overall performance, we have also analysed the performance of each forecasting model across various forecast horizons. Fig. 7 illustrates the point forecast accuracy metrics for each forecast horizon. While for shorter horizons, the performance of models are closer, there is a substantial gain in longer horizons using our proposed approach. For longer forecasting horizons, the proposed model retains high forecasting accuracy compared to other methods. Clearly, the proposed model not only remains accurate across the forecasting horizon, but is also more robust than other methods against increasing horizons.

Fig. 8 presents measures of the uncertainty of the models used in this study. Figure (8-A) shows the Winkler score for the 95% prediction intervals for each horizon. The proposed model provides more accurate prediction interval forecasts for all horizons and there is a substantial gain in using our model as forecast horizon increases. All benchmarks provide either more wider intervals and/or fail to cover the actual values inside the prediction interval.

In evaluating the probabilistic forecast accuracy of our model and its benchmark, we focus on the distributional characteristics of the generated forecasts evaluated by the corresponding percentile score and CRPS. Figure (8-B and C) presents the corresponding percentile score and CRPS aggregated for each forecasting horizon. The results show that the proposed model captures the density structure for consistently across all horizons better than the benchmark models. The difference is substantial for longer forecast horizons.

Figs. 7 and 8 indicate that the forecast accuracy of all models deteriorate as forecast horizon increases. Therefore, the further
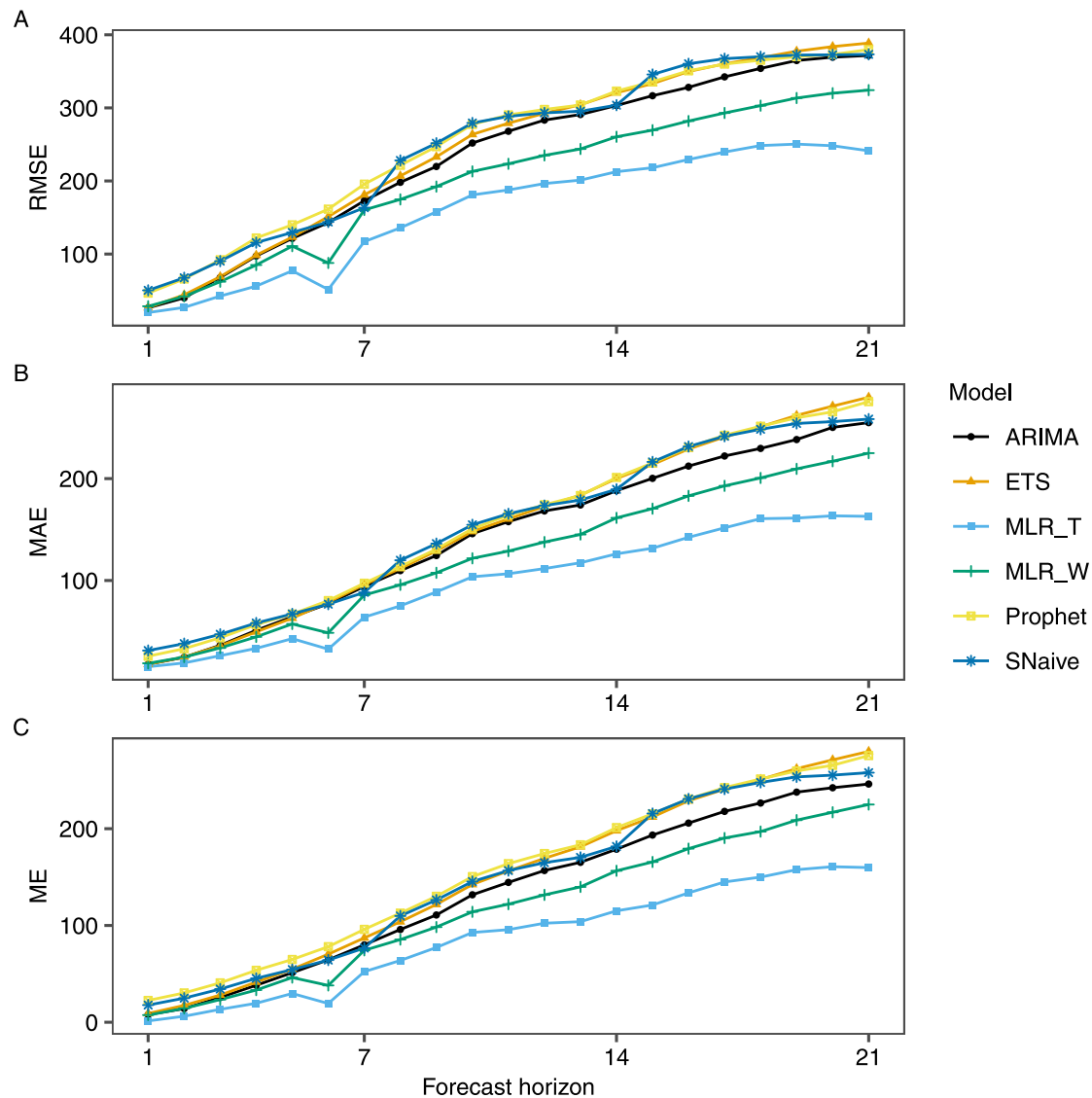
Fig. 7. Point forecast accuracy of models for each forecast horizon.

Table 1
Forecast performance evaluation.

Model	Accuracy measure					
	ME	RMSE	MAE	Winkler	Percentile	CRPS
Prophet	154.42	275.13	155.37	4465.83	73.10	145.61
SNaive	146.41	273.67	153.87	2628.72	69.14	137.07
ETS	148.89	271.99	153.28	3548.38	70.43	139.93
ARIMA	134.99	259.98	144.84	2942.25	65.34	129.72
MLR_W	117.52	222.59	124.30	2295.61	54.76	108.73
MLR_T	**86.70**	**178.06**	**97.00**	**583.25**	**39.71**	**78.72**

into the future a forecast is required, the forecast error becomes larger for all models.

Our results show that the proposed model not only outperforms benchmarks in terms of generating the correct value of future confirmed cases, but is also the most accurate when providing consistent information about uncertainty around confirmed cases. The model is very simple and can be understood by any manager, it also provides the entire probability distribution for confirmed cases for a given day at the local level. Our model offers the ability to capture a range of possibilities regarding confirmed cases for any given day that is not contained in the

point forecasts. This is very important for decision makers and planners in healthcare, as it helps them to assess the risk and make better decisions in planning. Further research attempts are required on how to use probabilistic forecasts and interpret output to inform decisions in the healthcare.

We should note that, if the focus of the decision makes is more on the short-term horizons ($h \leq 7$), then it is possible to create a MLR based model using the principles discussed in this paper that outperforms benchmarks substantially.

5. Conclusion

An accurate forecast of the daily number of positive confirmed cases is crucial for resource planning. Many models have been proposed in the literature to forecast confirmed cases ranging from time series to epidemic forecasting models.

In this study, we have created a simple and interpretable forecasting model that can be used to forecast positive confirmed cases at the local level. The proposed MLR model exploits the relationship between the confirmed case and the phone call data (NHS 111 calls), in addition to other patterns such as trend, the effect of weekends and Autoregressive lags of confirmed cases. We compare the performance of the model with ETS, ARIMA,
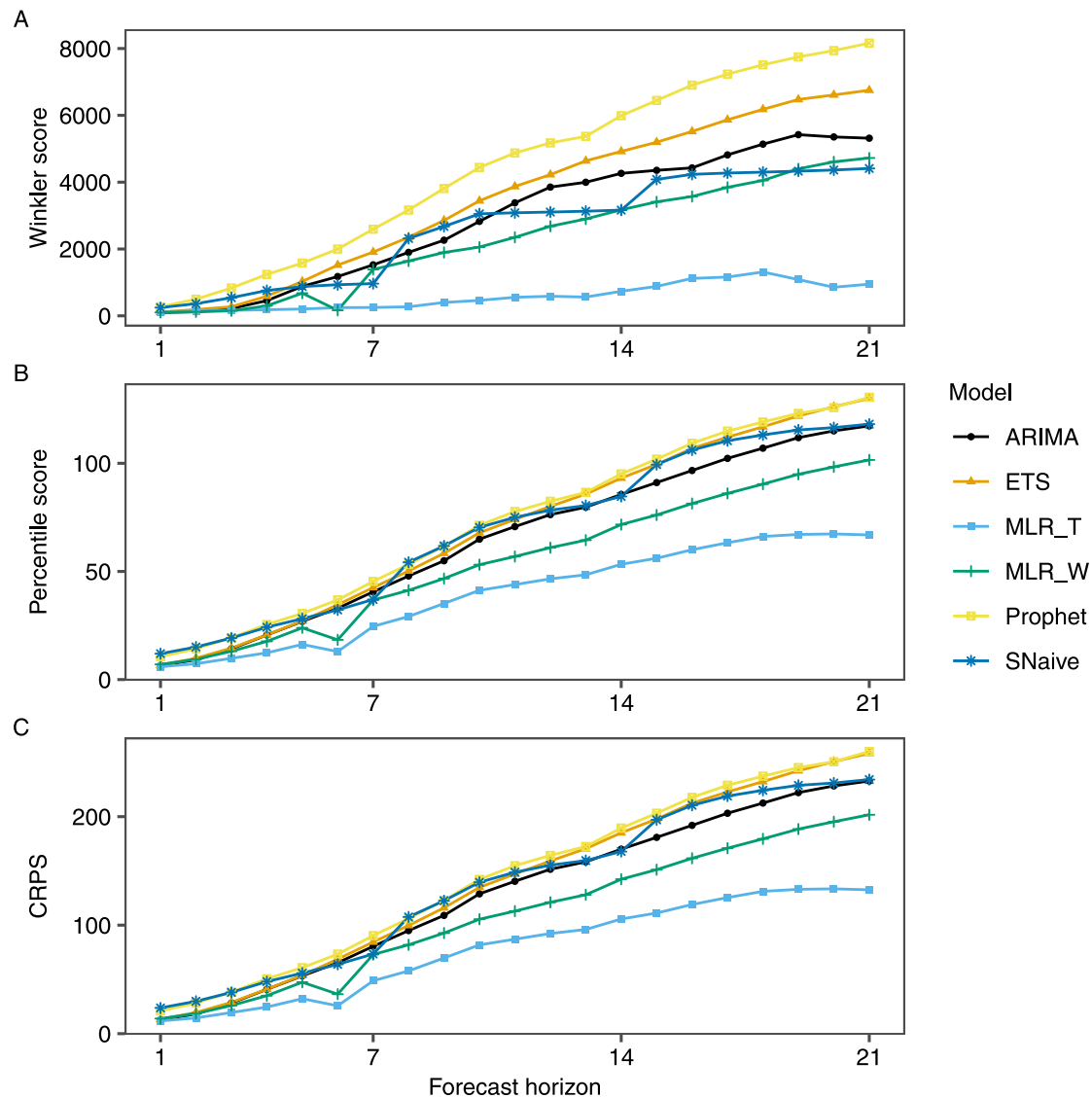
Fig. 8. Forecast uncertainty of models for each forecast horizon.

Seasonal Naive, Prophet and a MLR model without using phone call data using an empirical study.

These models are applied to the number of confirmed cases for a country council in England. In evaluating the performance of generated forecasts, we have used three point forecast error measures, one prediction interval accuracy measure and two probabilistic forecasting accuracy measures. Our analysis showed that the proposed model can provide accurate and reliable forecasts. It outperforms all benchmarks based on all accuracy measures considered in the study. We also provide evidence that using phone call data is an important predictor of COVID-19 confirmed cases and should be considered in forecasting models. We could propose that this might be due to the connection between phone calls to the health service and the dynamics related to COVID-19. The factors associated to being affected by the virus are clearly social, including deprivation, underlying conditions and belonging to ethnic minorities [42,43], and these factors are complex, dynamically entangled with mobility, economic activity, lock-downs and other government interventions. Phone calls to a centralised health service, such as the NHS, serve as an information junction where the chains of causality, related to social factors, are manifested. The effect of this source of information, renders possible the interesting performance of a relatively simple model.

In addition to generating traditional point forecasts, we also provide the probabilistic forecasts. Given the uncertainty involved in forecasting the number of COVID-19 confirmed cases, this is extremely important for decision makers as it highlights the uncertainty around a single value forecast. It can be used to manage risk in allocating resources in hospitals.

We have also used the proposed model to forecast the number of confirmed cases at the national level, England. The overall performance is very similar to the local level, the proposed model outperforms others. However, we have not provided the analysis here due to the space limit and the focus of the study at the local level, but it can be provided on request.

One of the limitations of this study is the access to more granular data such as age group, minority group, health condition, local regulations, knockdown, duration of knockdown, and so on. These might allow for a better explanation of changes in the number of confirmed cases. Further improvements of the proposed model can be achieved by incorporating additional information. Another limitation of the study is that the dataset does not contain hospital information such as COVID-19 admission and bed occupancy. This information allows for translating the forecast of COVID-19 confirmed cases into hospital admissions and bed capacity. This could be considered as an important research avenue.

CRediT authorship contribution statement

Bahman Rostami-Tabar: Conceptualization, Programming, Formal analysis, Model development, Writing - Original draft preparation. **Juan F. Rendon-Sanchez:** Conceptualization, Literature review, Writing - Original draft preparation.

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

References

[1] World Health Organisation, WHO Coronavirus disease (COVID-19) dashboard, 2020, https://covid19.who.int/.

[2] F.A. Binti Hamzah, C. Lau, H. Nazri, D.V. Ligot, G. Lee, C.L. Tan, M.K.B.M. Shaib, et al., Coronatracker: Worldwide Covid-19 outbreak data analysis and prediction, Bull World Health Organ 1 (32) (2020) http://dx.doi.org/10.2471/BLT.20.255695.

[3] Matjaž Perc, Nina Gorišek Miksić, Mitja Slavinec, Andraž Stožer, Forecasting Covid-19, Front. Phys. 8 (127) (2020).

[4] Marcel Salathe, Linus Bengtsson, Todd J. Bodnar, Devon D. Brewer, John S. Brownstein, Caroline Buckee, Ellsworth M. Campbell, et al., Digital epidemiology, PLoS Comput. Biol. 8 (7) (2012) 1–3, http://dx.doi.org/10.1371/journal.pcbi.1002616.

[5] Gitanjali R. Shinde, Asmita B. Kalamkar, Parikshit N. Mahalle, Nilanjan Dey, Jyotismita Chaki, Aboul Ella Hassanien, Forecasting models for coronavirus disease (Covid-19): A survey of the state-of-the-art, SN Comput. Sci. 1 (4) (2020) 1–15.

[6] John P.A. Ioannidis, Sally Cripps, Martin A. Tanner, Forecasting for Covid-19 has failed, Int. J. Forecast. (2020) http://dx.doi.org/10.1016/j.ijforecast.2020.08.004.

[7] George E.P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, Greta M. Ljung, Time Series Analysis: Forecasting and Control, John Wiley & Sons, 2015.

[8] James Durbin, Siem Jan Koopman, Time Series Analysis By State Space Methods, Vol. 38, Oxford University Press, 2012.

[9] Rob J. Hyndman, George Athanasopoulos, Forecasting: Principles and practice, 2020, OTexts. https://otexts.com/fpp3.

[10] Sean J. Taylor, Benjamin Letham, Forecasting at scale, Amer. Statist. 72 (1) (2018) 37–45.

[11] Fotios Petropoulos, Spyros Makridakis, Forecasting the novel coronavirus Covid-19, PLOS ONE 15 (3) (2020) 1–8, http://dx.doi.org/10.1371/journal.pone.0231236.

[12] Domenico Benvenuto, Marta Giovanetti, Lazzaro Vassallo, Silvia Angeletti, Massimo Ciccozzi, Application of the arima model on the covid-2019 epidemic dataset, Data Brief 29 (2020) 105340, http://dx.doi.org/10.1016/j.dib.2020.105340.

[13] Navid Feroze, Forecasting the patterns of Covid-19 and causal impacts of lockdown in top five affected countries using Bayesian structural time series models, Chaos Solitons Fractals 140 (2020) 110196, http://dx.doi.org/10.1016/j.chaos.2020.110196.

[14] Kay H. Brodersen, Fabian Gallusser, Jim Koehler, Nicolas Remy, Steven L. Scott, Inferring causal impact using Bayesian structural time-series models, Ann. Appl. Stat. 9 (2015) 247–274.

[15] Simon James Fong, Gloria Li, Nilanjan Dey, Rubén Gonzalez-Crespo, Enrique Herrera-Viedma, Finding an accurate early forecasting model from small dataset: A case of 2019-nCoV novel coronavirus outbreak, Int. J. Interact. Multimed. Artif. Intell. 6 (1) (2020b) 132, http://dx.doi.org/10.9781/ijimai.2020.02.002.

[16] Simon James Fong, Gloria Li, Nilanjan Dey, Rubén Gonzalez-Crespo, Enrique Herrera-Viedma, Composite Monte Carlo decision making under high uncertainty of novel coronavirus epidemic using hybridized deep learning and fuzzy rule induction, Appl. Soft Comput. 93 (2020a) 106282, http://dx.doi.org/10.1016/j.asoc.2020.106282.

[17] Muhammad Yousaf, Samiha Zahir, Muhammad Riaz, Sardar Muhammad Hussain, Kamal Shah, Statistical analysis of forecasting Covid-19 for upcoming month in Pakistan, Chaos Solitons Fractals 138 (2020) 109926, http://dx.doi.org/10.1016/j.chaos.2020.109926.

[18] Leila Moftakhar, S.E.I.F Mozhgan, Marziyeh Sadat Safe, Exponentially increasing trend of infected patients with Covid-19 in Iran: A comparison of neural network and arima forecasting models, Iran. J. Publ. Health 49 (2020) 92–100.

[19] Rajan Gupta, Saibal Kumar Pal, Trend analysis and forecasting of Covid-19 outbreak in India, 2020, medRxiv.

[20] Yonar Harun, Yonar Aynur, Agah Tekindal Mustafa, Tekindal Melike, Modeling and forecasting for the number of cases of the Covid-19 pandemic with the curve estimation models, the box-jenkins and exponential smoothing methods, EJMO 4 (2) (2020) 160–165, http://dx.doi.org/10.14744/ejmo.2020.28273.

[21] Barenya Bikash Hazarika, Deepak Gupta, Modelling and forecasting of Covid-19 spread using wavelet-coupled random vector functional link networks, Appl. Soft Comput. 96 (2020) 106626, http://dx.doi.org/10.1016/j.asoc.2020.106626.

[22] Anuradha Tomar, Neeraj Gupta, Prediction for the spread of Covid-19 in India and effectiveness of preventive measures, Sci. Total Environ. 728 (2020) 138762, http://dx.doi.org/10.1016/j.scitotenv.2020.138762.

[23] Vinay Kumar Reddy Chimmula, Lei Zhang, Time series forecasting of Covid-19 transmission in Canada using lstm networks, Chaos Solitons Fractals 135 (2020) 109864, http://dx.doi.org/10.1016/j.chaos.2020.109864.

[24] Gabriele Martelloni, Gianluca Martelloni, Modelling the downhill of the Sars-Cov-2 in Italy and a universal forecast of the epidemic in the world, Chaos Solitons Fractals 139 (2020) 110064, http://dx.doi.org/10.1016/j.chaos.2020.110064.

[25] Duccio Fanelli, Francesco Piazza, Analysis and forecast of Covid-19 spreading in China, Italy and France, Chaos Solitons Fractals 134 (2020) 109761, http://dx.doi.org/10.1016/j.chaos.2020.109761.

[26] Mohd Hafiz Mohd, Fatima Sulayman, Unravelling the myths of R0 in controlling the dynamics of Covid-19 outbreak: A modelling perspective, Chaos Solitons Fractals 138 (2020) 109943, http://dx.doi.org/10.1016/j.chaos.2020.109943.

[27] Khondoker Nazmoon Nabi, Forecasting Covid-19 pandemic: A data-driven analysis, Chaos Solitons Fractals 139 (2020) 110046, http://dx.doi.org/10.1016/j.chaos.2020.110046.

[28] Kankan Sarkar, Subhas Khajanchi, Juan J. Nieto, Modeling and forecasting the Covid-19 pandemic in India, Chaos Solitons Fractals 139 (2020) 110049, http://dx.doi.org/10.1016/j.chaos.2020.110049.

[29] Tridip Sardar, Sk Shahid Nadim, Sourav Rana, Joydev Chattopadhyay, Assessment of lockdown effect in some states and overall India: A predictive mathematical study on Covid-19 outbreak, Chaos Solitons Fractals 139 (2020) 110078, http://dx.doi.org/10.1016/j.chaos.2020.110078.

[30] Duygu Balcan, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J. Ramasco, Alessandro Vespignani, Multiscale mobility networks and the spatial spreading of infectious diseases, Proc. Natl. Acad. Sci. 106 (51) (2009) 21484–21489.

[31] Duygu Balcan, Bruno Gonçalves, Hao Hu, José J. Ramasco, Vittoria Colizza, Alessandro Vespignani, Modeling the spatial spread of infectious diseases: The global epidemic and mobility computational model, J. Comput. Sci. 1 (3) (2010) 132–145.

[32] The Gleam Project, Covid-19 modelling, United States, 2020, https://covid19.gleamproject.org/.

[33] Dianbo Liu, Leonardo Clemente, Canelle Poirier, Xiyu Ding, Matteo Chinazzi, Jessica T. Davis, Alessandro Vespignani, Mauricio Santillana, A machine learning methodology for real-time forecasting of the 2019-2020 Covid-19 outbreak using internet searches, news alerts, and estimates from mechanistic models, 2020, arXiv Preprint arXiv:2004.04019.

[34] Andres Hernandez-Matamoros, Hamido Fujita, Toshitaka Hayashi, Hector Perez-Meana, Forecasting of Covid19 per regions using arima models and polynomial functions, Appl. Soft Comput. 96 (2020) 106610, http://dx.doi.org/10.1016/j.asoc.2020.106610.

[35] Cleo Anastassopoulou, Lucia Russo, Athanasios Tsakris, Constantinos Siettos, Data-based analysis, modelling and forecasting of the Covid-19 outbreak, PLOS ONE 15 (3) (2020) 1–21.

[36] Clément Massonnaud, Jonathan Roux, Pascal Crépey, COVID-19: Forecasting short term hospital needs in France, 2020, medRxiv.

[37] Gary E. Weissman, Andrew Crane-Droesch, Corey Chivers, ThaiBinh Luong, Asaf Hanish, Michael Z. Levy, Jason Lubken, et al., Locally informed simulation to predict hospital capacity needs during the Covid-19 pandemic, Ann. Intern. Med. (2020).

[38] Mitchell O'Hara-Wild, Rob Hyndman, Earo Wang, Gabriel Caceres, Fable: Forecasting models for tidy time series, 2020.

[39] Rob J. Hyndman, Anne B. Koehler, Another look at measures of forecast accuracy, Int. J. Forecast. 22 (4) (2006) 679–688.

[40] Robert L. Winkler, A decision-theoretic approach to interval estimation, J. Amer. Statist. Assoc. 67 (337) (1972) 187–191.

[41] Tao Hong, Pierre Pinson, Shu Fan, Hamidreza Zareipour, Alberto Troccoli, Rob J. Hyndman, Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond, Int. J. Forecast. 32 (2016) 896–913.

[42] Thomas Baumer, Emily Phillips, Amrit Dhadda, Tamas Szakmany, Epidemiology of the first wave of Covid-19 Icu admissions in south Wales–the interplay between ethnicity and deprivation, Front. Med. 7 (2020).

[43] Sarah Caul, Deaths involving Covid-19 by local area and socioeconomic deprivation: Deaths occurring between 1 march and 31 2020, 2020, https://backup.ons.gov.uk/wp-content/uploads/sites/3/2020/08/Deaths-involving-COVID-19-by-local-area-and-socioeconomic-deprivation-deaths-occurring-between-1-March-and-31-.pdf.