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

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
A time series-based statistical approach for outbreak spread forecasting: Application of COVID-19 in Greece

Christos Katris

1

Athens University of Economics and Business, Department of Accounting and Finance, Greece

ARTICLE INFO

Keywords:
Time series forecasting
Classical time series models
Machine learning approaches
Forecast combinations
tSIR epidemiological model

ABSTRACT

The aim of this paper is the generation of a time-series based statistical data-driven procedure in order to track an outbreak. At first are used univariate time series models in order to predict the evolution of the reported cases. Moreover, are considered combinations of the models in order to provide more accurate and robust results. Additionally, statistical probability distributions are considered in order to generate future scenarios. Final step is the build and use of an epidemiological model (tSIR) and the calculation of an epidemiological ratio ($R_0$) for estimating the termination of the outbreak. The time series models include Exponential Smoothing and ARIMA approaches from the classical models, also Feed-Forward Artificial Neural Networks and Multivariate Adaptive Regression Splines from the machine learning toolbox. Combinations include simple mean, Newbolt-Granger and Bates-Granger approaches. Finally, the tSIR model and the $R_0$ ratio are used for estimating the spread and the reversion of the pandemic. The suggested procedure is used to track the COVID-19 epidemic in Greece. This epidemic has appeared in China in December 2019 and has been widespread since then to all over the world. Greece is the center of this empirical study as is considered an early successful paradigm of resistance against the virus.

1. Introduction

Coronavirus disease 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The virus appeared in China in December 2019 and has since spread globally. There are many victims of the virus (over 750,000) – and much more reported cases (over 20000000) until August 14 of 2020. The spread of the virus led World Health Organization to characterize it as a pandemic in 11 March 2020. Accurate modeling and future forecast of daily number of confirmed cases can help the treatment system in providing services for the new patients. Certain statistical models from the time series domain can help in this direction using only historical data of the daily confirmed cases. In this work are used classical time series models: Exponential Smoothing models and Auto Regressive Integrated Moving Average (ARIMA), machine learning approaches: Multivariate Adaptive Regression Splines (MARS) and Feed-Forward Artificial Neural Networks (ANN) and combinations of the above classes with the aim to gain - if possible - additional accuracy. The best forecasting model is decided through the Mean Absolute Percentage Error (MAPE). Furthermore, different probability distributions are considered for the modeling of reported cases data and also are evaluated through Bayes Information Criterion. Using the selected approaches from forecasting models and also from probability distribution fitting, there are considering and analyzing different scenarios according to the severity of the future situation. Finally, the tSIR epidemiological model is used both for estimation of certain measures which are the expected reported rate and the estimation of the termination of the pandemic in each country. Additionally, is also computed the $R_0$. These metrics allow the comparison of the spread between different countries.

So far, in order to forecast the spread of COVID-19 there have been done some attempts such as (Fanelli and Piazza, 2020) where they forecast the spread epidemics in France, Italy and China using a SIRD epidemic model, Wu et al. (2020) used a SIR epidemiological model to simulate the dynamics of the disease across major cities in China, Li et al. (2020) use a transmission model for studying early transmission dynamics in Wuhan, Zhao et al. (2020) made a preliminary estimation of the basic reproduction number of novel coronavirus in China at the early stages of the outbreak by using an epidemiological approach, Anastassopoulou et al. (2020) use the SIRD model in order to provide estimates of the basic epidemiological parameters and to forecast the
evolution of COVID-19. Additionally, Roosa et al. (2020) use a generalized logistic growth model, the Richards growth model and a wave growth model in order to generate forecasts of the reported cases in Guangdong and Zhejiang in China and also Roosa et al. (2020) provide real-time forecasts for COVID-19 in China by using three phenomenological models. Ahmadi, Fadai, Shirani, and Rahmani (2020) are using epidemic projection models (logistic growth, Gompertz growth, von Bertalanffy and Cubic Polynomial least squared error (LSE) method) in order to calculate basic reproduction number and predict the epidemic trend in Iran. Petropoulos and Makridakis (2020) use Exponential smoothing family models in order to predict confirmed cases of COVID-19. Finally, some machine learning approaches have been applied such as these of Hu, Qiyang, Shudi, Li, and Momiao (2020) who built an autoencoder in order to forecast the spread of the virus in China and of Al-qaness et al. (2020) who propose an ANFIS method in order to shown that such methods prevail compared to classical models accord linearity which cannot be captured accurately from classical time se.

2. Models

The use of classical time series models is advantageous in many ways. Some of them are mainly that historical values often display a linear pattern and these models can capture such structure. Also, these models have a clear interpretation of the coefficients and can offer insights about the functionality of the system and are relatively easy to be computed as they have already used extensively for many years.

On the other hand, lies the question why someone should use machine learning approaches. These are relatively modern methods and their use for time series prediction is quite recent. The most obvious point for their use is when there are signs of existence of non-linearity (according to statistical tests or maybe potential undetected non-linearity) which cannot be captured accurately from classical time series models. Additionally, there are applications in many fields where is shown that such methods prevail compared to classical models according to forecasting accuracy. Finally, the models of this field constitute a totally different approach in time series prediction and their use in conjunction with classical models could offer gains in forecasting accuracy.

2.1. Classical time series models

The first rational approach in forecasting cases is the use of classical time series models. There are used the Autoregressive Integrated Moving Average (ARIMA) model, the Holt’s model (DES), the Holt-Winters model (TES) and simple Exponential Smoothing model as benchmark.

Exponential Smoothing (ES) models (Hyndman & Athanasopoulos, 2018) use a weighted moving average with weights that decrease exponentially. The basic idea is that the weights decay exponentially as the observations getting older. The intention is to weigh heavier the more recent data. The simpler case is the simple Exponential Smoothing (SES) which is suitable for data with no clear trend or seasonality. The SES model lies between the average forecast of previous values and the naive method forecast (where the forecast for the future is simply equal to the last observed value of the series). Let the previous observed values of the series be \( y_1, y_2, \ldots, y_t \). Then the forecast for \( y_{t+1} \) is \( \hat{y}_{t+1} = \alpha y_t + (1-\alpha) \hat{y}_t \) where \( 0 \leq \alpha \leq 1 \) is the smoothing parameter. This model leads to flat forecasts (all forecasts after the last observation take the same value). This simpler case is used as benchmark in this study. This SES model has been extended by Holt (1957) in order to allow a trend in data (DES model). This model is presented by one equation for forecast and two smoothing equations for the level and the trend respectively. The forecasts are \( \hat{y}_{t+h} = l_t + b_t h \), where the level at time \( t \) is \( l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1}) \) with \( 0 \leq \alpha \leq 1 \) to be the smoothing parameter for the level and the trend at time is \( b_t = \beta (l_t - l_{t-1}) + (1-\beta) b_{t-1} \) with \( 0 \leq \beta \leq 1 \) to be the smoothing parameter for the trend. The h-step-ahead forecast is equal to the last estimated level plus h times the last estimated trend value.

The DES model has been expanded by Holt and Winters in order to capture seasonality (TES model). There are two alternatives of the method, the additive method which is preferred when the seasonal variations are almost constant through the series and the multiplicative method when the seasonal variations change proportional to the level of the series. In this work, is used the additive method. The form of this model includes one equation for forecasting \( \hat{y}_{t+h} = l_t + b_t h + s_t h + m_k (t + 1) \) and 3 smoothing equations for the level \( l_t = \alpha (y_t - s_t m_t) + (1-\alpha)(l_{t-1} + b_{t-1}) \), the trend \( b_t = \beta (l_t - l_{t-1}) + (1-\beta) b_{t-1} \) and the seasonality \( s_t = \gamma (y_t - l_t - b_t) + (1-\gamma) s_{t-1} \) respectively.

The data in this study are daily and from one season and naturally annual seasonality can be ignored. However, there is possibility that the data from each country display cyclic behavior and for this reason the extension of Holt-Winters model which includes seasonality is examined.

The implementation of Exponential Smoothing type models is performed using R package forecast (Hyndman & Khandakar, 2007).

ARIMA model introduced by Box and Jenkins and is analyzed in detail in Box et al. (2015) and is widely used for modeling and analyzing time series in many fields. Examples include: the prediction and monitoring of the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore (Earnest et al., 2005), the prediction of Dengue Haemorrhagic Fever Cases in Southern Thailand (Promprou et al., 2018) use a weighted moving average with weights that decrease exponentially as the observations getting older. The intention is to weigh heavier the more recent data. The simpler case is the simple Exponential Smoothing (SES) which is suitable for data with no clear trend or seasonality. The SES model lies between the average forecast of previous values and the naive method forecast (where the forecast for the future is simply equal to the last observed value of the series). Let the previous observed values of the series be \( y_1, y_2, \ldots, y_t \). Then the forecast for \( y_{t+1} \) is \( \hat{y}_{t+1} = \alpha y_t + (1-\alpha) \hat{y}_t \) where \( 0 \leq \alpha \leq 1 \) is the smoothing parameter. This model leads to flat forecasts (all forecasts after the last observation take the same value). This simpler case is used as benchmark in this study. This SES model has been extended by Holt (1957) in order to allow a trend in data (DES model). This model is presented by one equation for forecast and two smoothing equations for the level and the trend respectively. The forecasts are \( \hat{y}_{t+h} = l_t + b_t h \), where the level at time \( t \) is \( l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1}) \) with \( 0 \leq \alpha \leq 1 \) to be the smoothing parameter for the level and the trend at time is \( b_t = \beta (l_t - l_{t-1}) + (1-\beta) b_{t-1} \) with \( 0 \leq \beta \leq 1 \) to be the smoothing parameter for the trend. The h-step-ahead forecast is equal to the last estimated level plus h times the last estimated trend value.

The DES model has been expanded by Holt and Winters in order to capture seasonality (TES model). There are two alternatives of the method, the additive method which is preferred when the seasonal variations are almost constant through the series and the multiplicative method when the seasonal variations change proportional to the level of the series. In this work, is used the additive method. The form of this model includes one equation for forecasting \( \hat{y}_{t+h} = l_t + b_t h + s_t h + m_k (t + 1) \) and 3 smoothing equations for the level \( l_t = \alpha (y_t - s_t m_t) + (1-\alpha)(l_{t-1} + b_{t-1}) \), the trend \( b_t = \beta (l_t - l_{t-1}) + (1-\beta) b_{t-1} \) and the seasonality \( s_t = \gamma (y_t - l_t - b_t) + (1-\gamma) s_{t-1} \) respectively.

The data in this study are daily and from one season and naturally annual seasonality can be ignored. However, there is possibility that the data from each country display cyclic behavior and for this reason the extension of Holt-Winters model which includes seasonality is examined.

The implementation of Exponential Smoothing type models is performed using R package forecast (Hyndman & Khandakar, 2007).

ARIMA model introduced by Box and Jenkins and is analyzed in detail in Box et al. (2015) and is widely used for modeling and analyzing time series in many fields. Examples include: the prediction and monitoring of the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore (Earnest et al., 2005), the prediction of avian influenza H5N1 (Kane et al., 2014), the prediction of Dengue Haemorrhagic Fever Cases in Southern Thailand (Promprou et al., 2018). For this situation, the most recent example is the use of this model for forecasting confirmed cases of COVID-19 in different countries (Deleshi et al., 2020). It is well-known that ARIMA models frequently outperform more sophisticated structural models in terms of short-run forecasting ability (Meyer et al., 1998). ARIMA models can be noted as ARIMA (p,d,q), where p is the order of the autoregressive part, d is the order of differencing and q the moving average part. The first step is to use differencing in order to make the series stationary. Considering that d is the number of differences which are needed to make the series stationary, i.e. \( d \Delta^d Y_t = (1-L)^d Y_t = Y_t - Y_{t-d} \), then the model which is fitted to the original series (ARIMA (p,d,q)) has the form:

\[
A(L) (1-L)^d Y_t = \delta + \Theta(L) \varepsilon_t,
\]

where \( A(L) = 1 - \alpha_1 L - \ldots - \alpha_p L^p \) and \( \Theta(L) = 1 - \theta_1 L - \ldots - \theta_q L^q \).

More details about the model can be found in Box and Jenkins (1976). The order of the model is decided based on the BIC criterion and the estimation of the parameters is performed using the forecast package of R software (Hyndman & Khandakar, 2007).
2.2. Machine learning approaches

These models are relatively modern approaches, compared to the classical time series models. The first considered machine learning technique is the Artificial Neural Networks (ANN) approach which has been successfully applied to many different areas (e.g. Lippmann, 1987; Zhang, Patuwo, & Hu, 1998). ANN forecasting models for time series use a set of the k most recent values for predicting the next one. There have been a number of different architectures for ANNs, but in the current work the Feed-Forward fully connected Neural Networks are used in this work. This is one of the most popular architectures of ANN models.

After the decision of the architecture for the ANN, the input variables, the number of hidden layers and the number of nodes for each layer have to be decided. As input, we consider one and two previous observations which are the number of reported cases of the previous day (or of 2 previous days). We keep the input up to two nodes for simplicity as the complexity increases a lot with additional input and we want to avoid the risk of overfitting (Takagi, Livingstone, & Luik, 1995). Empirical research has shown that one hidden layer is sufficient in most cases (Haykin, 2010). The data are normalized using a z-score normalization with the mean and standard deviation of the training sample. Each layer is fully connected to the next one and the activation function used in the hidden layer is the sigmoid

\[ S(t) = \frac{1}{1 + e^{-x}} \]

Additionally, a linear function is used in the output layer in order to transform the previous inputs to final outputs. The output layer includes simply the forecasts, which then are transformed back to original values. The training of the network has been done with the adaptive back-propagation with momentum method, where the weights of the connections in the neural network are estimated using the adaptive gradient descent optimization algorithm (Haykin, 2010) and the implementation is performed through the R package AMORE (Limas et al., 2014).

The method of multivariate adaptive regression splines (MARS) introduced by Friedman in 1991 (Friedman, 1991) and the extension of MARS to a time series context is discussed in the paper of (Lewis & Stevens, 1991). It is a non-parametric regression which extends linear models by automatic modeling of non-linearities and interactions between variables. The system which generated the data is supposed to have the form:

\[ y = f(x) + \epsilon \]

where y is the dependent variable, in our case \( y = [y_1, y_2, \ldots, y_p]^T \) is a vector of predictors which are the lagged time series variables and \( \epsilon \) is an additive stochastic component with zero mean and finite variance. In MARS method, nonparametric models are developed locally, in certain sub-regions of the data and by this way increases flexibility of a global modeling approach to the data. The detection of the optimal number of sub-regions and the fitting of different regression lines for each sub-region is the key for the achievement of flexibility and non-linearity. For each predictor variable, the method detects ‘knots’ (i.e. breakpoints) which define the intervals where the regression slope changes. The predictor variables which are included in the final model and their corresponding ‘knots’ are determined through an intensive search procedure. The term of hinge function is central and has the form \( \max(0, x-c) \) or \( \max(0, c-x) \), where c is a constant (a ‘knot’).

The built of the model have 2 phases: the forward and the backward pass. In the forward pass the model starts with an intercept term and variables are added. At each step the model finds 2 basis functions which give the minimum reduction in sum of squares residual error. A new basis function is a multiplication of an existing term in the model and a new hinge function. The forward procedure stops when the change in residual error is too small to continue or the maximum number of terms has been reached. This step results often in an overfit model. The backward pass has the scope of build a model with better predictive ability (less adhered to the training set). In order to achieve this, the model is pruned in this step, the less effective terms are detected and removed until the “best” model is found according to

\[ GCV = \frac{RSS}{N - \text{dof} - p} \]

criterion. (RSS is the Residual Sum of Squares, N is the number of observations and \#Bf Par = (\text{eq}(\text{MARSterms}) + \text{penalty}) / (\text{eq}(\text{MARSterms}) - 1) \) is the effective number of parameters. The penalty term is usually 2 or 3. In this work, MARS model with lagged variables as predictors is implemented in the earth package of R statistical software (Milborrow, 2019). In this work, lagged variables with 1 and 2 lagged values are considered as inputs to construct models and we select the one with minimum RMSE.

2.3. Combinations

The following different combination schemes are used in order to improve further the forecasting accuracy of the models:

| Method           | Formula               |
|------------------|-----------------------|
| Bates/Granger    | \[ w_m = \frac{\sigma^2(i)}{\sum_{i=1}^{n} \sigma^2(j)} \] where \( \sigma^2 \) is estimated mean square of the i prediction model and the forecast combination is \( \hat{y}_t = f_w^{|w_m|} \) |
| Newbold/Granger  | \[ w_m = \frac{\epsilon w - 1}{\epsilon w} \] a constrained minimization of the mean squared prediction error with normalization condition \( \epsilon w - 1 \) and the forecast combination is \( \hat{y}_t = f_w^{\mid w_m \mid} \) |
| Simple Average   | \( \hat{y}_t = \frac{1}{N} \sum_{i=1}^{N} y_t^i \) |

Details about the formulas and the combinations can be found at forecasting package description: https://cran.r-project.org/web/packages/ForecastComb/ForecastComb.pdf

The aim is to further enhance the forecasting accuracy of individual approaches and also to provide a more robust solution in forecasting an outbreak. All models are implemented in R package ForecastComb (see Weiss et al., 2018). The description of the forecasting combinations and their corresponding formulas can be found also in Weiss et al (2018).

2.4. tsIR epidemiological model and R0 metric

One of the simplest and powerful mathematical models of infectious diseases is the Susceptible-Infected-Recovered (SIR) model (Anderson & May, 1991) and many other models are derivatives of this basic form.

The three variables of the model which represent the number of individuals in each situation are: the number of susceptible (S), the number of infectious (I) and the number of recovered or deceased (or immune) (R). In an epidemic often transitions from one situation to another is often much faster than the dynamics of birth and death (demographic processes), therefore is frequent to be omitted. However, following (Becker & Grenfell, 2017) the basic assumptions of the SIR model are: a well-mixed population, demographic characteristics (i.e. births (B) and deaths (D)) and infectious disease specific properties (i.e. contact rate, β, and infectious period, inverse γ, for a single pathogen. Other characteristics (thus, relevant symbols) of the model are that the transmission coefficient \( \beta \) varies seasonally (especially for the prototypical example, measles), \( \lambda = \frac{B}{N} \) is the force of infection and the immunity post-infection is life-long (although this assumption can be relaxed via a Susceptible-Infected-Recovered-Susceptible (SIRS) model).

The equations below describe the deterministic expression of the model:

\[ \frac{dS}{dt} = B - \frac{\lambda S}{N} - \mu S \]
\[ \frac{dI}{dt} = \frac{\lambda S}{N} - \gamma I - \mu I \]
\[ \frac{dR}{dt} = \gamma I - \mu R \]
Although simple, the above model is mathematically and statistically difficult to calibrate for time series data (Caudron et al., 2015; Takahashi et al., 2016) and the fitting of such models to time-series data is not trivial for some reasons (Becker & Grenfell, 2017). The SIR model for time series (tsIR) has been presented in (Finkenstadt & Grenfell, 2000). The implementation of the model in an R package named tsIR was introduced in (Becker & Grenfell, 2017).

Furthermore, in epidemiology $R_0$ is the basic reproduction number which represents the expected number of cases directly generated by one case in a population where all individuals are susceptible to infection. Further details about $R_0$ can be found in (Van den Driessche & Watmough, 2008). There are many methods of estimation methods for this number in epidemics, and the implementation of many estimation methods can be found on R package R0 (Boelle et al., 2015). In this work we use the Exponential Growth method (Wallinga & Lipsitch, 2007).

3. Suggested procedure

The main scope of this paper is to suggest a time-series based statistical data-driven approach for the handling of an outbreak situation. In the core of the method are the time series models in order to achieve accurate prediction, are also used statistical probability distributions in order to generate alternative scenarios for the future of the spread of the disease and finally is used the tsIR epidemiological model and $R_0$ epidemiological rate in order to track the evolution and also to estimate the termination of the outbreak. The steps of the suggested procedure are presented:

I. Forecasting Time Series of reported cases using Individual Models:
1. Fit Exponential Smoothing model
2. Fit Holt model
3. Fit Holt-Winters model
4. Fit ARIMA model (normal innovations)
5. Fit ARIMA model (skewed normal innovations)
6. Fit ANN Model
7. Fit MARS Model

II. Ranking Models according to MAPE Criterion

III. Forecasting Time Series of reported cases using Combinations of Models
1. Fit Bates/Granger Combination (of 2 and 3 models)
2. Fit Newbold/Granger Combination (of 2 and 3 models)
3. Fit Simple Mean Combination (of 2 and 3 models)

IV. Select the best forecasting approach according to MAPE Criterion

V. Fit probability distributions to the data
1. Fit Normal Distribution
2. Fit Log-Normal Distribution
3. Fit Exponential Distribution
4. Fit Pareto distribution

VI. Decide best fitting of a probability distribution according to AIC Criterion

VII. Create scenarios based on the selected models from steps (IV) and (VI)
1. good (25% quantile)
2. expected (forecast),
3. slightly negative (70% quantile),
4. negative (80% quantile)
5. bad (90% quantile)

VIII. Fit a tsIR epidemiological model to the data
1. Calculate Expected Reporting Rate
2. Estimate Expected Cycle Length

IX. Calculate the $R_0$ ratio

X. Estimate the reversion of the pandemic

It is clear that the above procedure has flexibility in each step as it does not recommend standard time series model (step I), specific probability distributions (step V), specific criteria for their evaluation (in steps II, IV and VI) and specific scenarios (step VII). The above selections depend on the researcher and the specific data set of the problem. For example, there could be a selection of specific forecasting models and/or probability distributions if previous studies suggest such a choice. Additionally, there is no global choice for criteria of evaluation: e.g. for some applications and/or researchers could be more important the more clear interpretation of the comparison of the models, while in other cases there are considered as more important terms criteria such as heavier error penalty in case of underestimation etc. Finally, the scenario generation is not restrictive: i.e. one can generate many alternative scenarios (not just 5) for a specific problem based on different quantities of the selected distribution.

4. Data analysis

4.1. Overview

4.1.1. Data

The data are publicly available from the European Centre for Disease Prevention and Control and are downloaded from the https://ourworldindata.org/coronavirus-source-data site. The data covers the time period from the beginning of the reported cases until August 14 of 2020.

4.1.2. Descriptive statistics and statistical tests

Table 1 displays descriptive statistics for 13 countries and for the whole World. The second column mention the total reported cases and the third the total deaths. The next two columns report the cases and the deaths in average for the last 10 days. An initial screening of the countries can be made, especially in the comparison with the whole world. For example, USA displays over 1150 (almost 18% of the whole world) and UK over 77 daily deaths in the last 10 days. On the other hand, all the other countries under examination display under 20 daily deaths with Spain to display almost 18 deaths and Italy to display only 6.5 daily deaths in the last 10 days. These 2 countries exhibited serious problem related to covid-19 previously but now seem more ‘robust’ against the virus. Finally, there are countries (Greece, Switzerland, Austria, Finland) with 2 or less daily deaths.

Additionally, for the number of new cases, we examine their normality using Shapiro-Wilk test and the potential existence of non-linearity using the White Neural-Network test (Lee et al., 1993). The results are displayed in Table 2.

From Table 1, we can see clearly that USA display the largest problem totally and the last 10 days. In Europe UK and Spain are the most negative occasions with almost 78 and 18 deaths daily the last 10 days, while France and Turkey have over 10 deaths daily at the same period.

| Country       | Total Cases | Total Deaths | Mean Cases | Mean Deaths |
|---------------|-------------|--------------|------------|-------------|
| Greece        | 6.381       | 221          | 164.4      | 1.714       |
| Italy         | 252.235     | 35.231       | 400.6      | 6,500       |
| Spain         | 337.334     | 28.605       | 4.931,4    | 18,167      |
| Germany       | 221.413     | 9.225        | 1.013,2    | 6,900       |
| Switzerland   | 37.312      | 1.714        | 178,5      | 1,800       |
| Austria       | 22.730      | 725          | 138,9      | 1,167       |
| Finland       | 7.685       | 333          | 21,7       | 2,000       |
| Sweden        | 83.852      | 5.776        | 355,0      | 4,710       |
| Turkey        | 245.635     | 5.912        | 1,178,4    | 16,500      |
| Netherlands   | 61.149      | 6.156        | 573,4      | 2,667       |
| UK            | 313.798     | 46.706       | 908,3      | 77,714      |
| USA           | 5.248.242   | 333.110      | 53.468,0   | 1,170,700   |
| France        | 209.365     | 30.388       | 2,258,8    | 11,750      |
| World         | 20.900.777  | 759.358      | 263.046.6  | 6,562,300   |

1 Last 10 days (August 5-August 14) – (For Spain until August 13).
4.3. Application of the method

Deviation from normality can be assumed at 1% level. This indicates whether a linear and/or normal distribution assumption can be rejected at 1% level and for some of the presented countries (but not worldwide) the assumption of linearity can exist. The non-linearity in many cases strengthens the choice for models from the machine learning domain.

4.3.3. Fit probability distributions to the data

In order to fit the distributions we select the most accurate in order to describe data according to Bayes Information Criterion (BIC). In order to fit the distributions is used the R package fitdistrplus (Delignette-Muller et al., 2019). Table 7 displays the results, i.e. the 1st column shows the name of the probability distribution, the 2nd column displays the parameters for each model and the 3rd column the value of BIC. The selected model is the Log-normal distribution due to the lowest value of BIC.

The forecasting approaches are evaluated the last seven days and the evaluation metric is the Mean Absolute Percentage Error (MAPE) of these days. The results are shown in Table 4 for the individual models.

It is observed from Table 4 that the best 3 models are Holt (DES model), ARIMA (normal innovations) and ANN with 1 input node. With these models we create combinations. The results are displayed in Table 5. Our choice is to use the NG combination of Holt, ARIMA (with Normal innovations) and ANN (with 1 input node) models.

4.3.2. The effect of seasonality

As is already mentioned in Section 2.1 the series could display cyclic behavior and for this reason models which take into account seasonality maybe could give more accurate results. Given that the Holt model is the most accurate, the Holt-Winters model extension (TES model) is examined. In order to check the effect of seasonality and further improve the forecasts, is applied the Holt-Winters model (TES model). It is unclear the period which should be used, but is applied an iterative procedure, where periods from 2 to 12 are considered and is selected the one with the minimum MAPE criterion. The selected frequency is nine as is gives the minimum MAPE with 0.025 which prevails over the other individual models. Fig. 1 is following and displays the results of this model selection:

Next, we consider the combination of Newbolt-Granger (NG combination) of TES, ARIMA (normal innovations) and ANN (with 1 input node) models. This combination is compared to the NG combination of DES, ARIMA (normal innovations) and ANN (with 1 input node) models. Table 6 displays the results of this comparison.

We observe that the consideration of a model with seasonality can add forecasting accuracy.

4.3.3. Fit probability distributions to the data

In order to build alternative scenarios for the evolution of the series, we are fitting some probability distributions to the data, by using the mean value and the standard deviation of the last 10 days to compute parameters. We compare and select probability distribution among: Normal, Exponential, Log-normal and Pareto models. From these probability distributions, we select the most accurate in order to describe data according to Bayes Information Criterion (BIC). In order to fit the distributions is used the R package fitdistrplus (Delignette-Muller et al., 2019). Table 7 displays the results, i.e. the 1st column shows the name of the probability distribution, the 2nd column displays the parameters for each model and the 3rd column the value of BIC. The selected model is the Log-normal distribution due to the lowest value of BIC.

Table 2

| Country     | Normality (Shapiro-Wilk) Statistic | p-value | Non-Linearity (White Test) Statistic | p-value |
|-------------|-----------------------------------|---------|--------------------------------------|---------|
| Greece      | 0.7569                            | <0.001  | 1.0685                               | 0.586   |
| Italy       | 0.7652                            | <0.001  | 2.7665                               | 0.252   |
| Spain       | 0.7970                            | <0.001  | 10.1190                              | 0.006   |
| Germany     | 0.7079                            | <0.001  | 7.4729                               | 0.024   |
| Switzerland | 0.6822                            | <0.001  | 8.9909                               | 0.014   |
| Austria     | 0.6517                            | <0.001  | 36.2820                              | <0.001  |
| Finland     | 0.8290                            | <0.001  | 40.9580                              | <0.001  |
| Sweden      | 0.8277                            | <0.001  | 19.4260                              | <0.001  |
| Turkey      | 0.8166                            | <0.001  | 4.0226                               | 0.134   |
| Netherlands | 0.8231                            | <0.001  | 6.7370                               | 0.034   |
| UK          | 0.8469                            | <0.001  | 8.8544                               | 0.012   |
| USA         | 0.9309                            | <0.001  | 4.5884                               | 0.101   |
| France      | 0.8083                            | <0.001  | 9.8963                               | 0.007   |
| World       | 0.8944                            | <0.001  | 8.2476                               | 0.016   |

Table 3

| Stationarity ADF Test | Autocorrelation Ljung-Box test | Non-Linearity White NN test | Normality Shapiro-Wilk |
|-----------------------|-------------------------------|-----------------------------|------------------------|
| Greece                | 1.2545 (0.99)                 | 91.1990 (<0.001)            | 1.0685 (0.586)         | 286.2400 (<0.001) |

Table 4

| MODEL                              | MAPE  |
|------------------------------------|-------|
| Exponential Smoothing (Benchmark)  | 0.112 |
| Holt (Additive)                    | 0.028 |
| ARIMA (normal innovations)         | 0.044 |
| ARIMA (skewed normal innovations)  | 0.072 |
| MARS                               | 0.063 |
| ANN (1 input node)                 | 0.046 (n = 5) |
| ANN (2 input nodes)                | 0.065 (n = 1) |
4.3.4. Scenario generation

Using the selected forecasting model (NG combination of TES, ARIMA and ANN with 1 input node) and the selected probability distribution (Log-normal), are generated 5 scenarios for the evolution of the reported cases: good (25% quantile), expected (forecasts), slightly negative (70%), negative (80% quantile) and bad (90% quantile). These scenarios are shown in Table 8 which displays the cumulative cases for each scenario. The construction of alternative scenarios is very important and helpful for monitoring the evolution of the spread of the outbreak, for evaluating the fighting strategy against its spread and also adjusting it if needed. Also, these alternative scenarios for the future are displayed in Fig. 2.

| Table 5 | Evaluation of Combinations. |
|---------|----------------------------|
| MODEL                           | MAPE   |
| Simple Mean Combination (2 models) | 0.0343 |
| NG Combination (2 models)       | 0.0179 |
| BG Combination (2 models)       | 0.0305 |
| Simple Mean Combination (3 models) | 0.0383 |
| NG Combination (3 models)       | 0.0828 |
| BG Combination (3 models)       | 0.0313 |

| Table 7 | Comparison of Probability Distributions. |
|---------|-----------------------------------------|
| MODEL  | Parameters | BIC |
| Normal | m = 38.20, sd = 43.40                    | 1743.516 |
| Exponential | λ = 0.0262      | 1555.856 |
| Log-normal | m = 3.140, sd = 1.122            | 1505.751 |
| Pareto | shape = 6.5625, scale = 213.1467      | 1557.730 |

4.3.5. Fit a tSIR epidemiological model to the data and compute R₀

In order to implement such a model we need additional data related with Population and birth rates (crude per 1000 persons) of countries. The data are coming from the World Bank. The latest data are for 2018. Table 9 displays the expected reporting rate, the estimated cycle length of the series (by finding frequency using the R package forecast) and the Basic Reproduction number using the Exponential Growth method with R package R0.

The expected reporting rate is the average of the last 10 days of the data. We let 4 weeks (28 days) out of the sample ("burn" the data) to obtain more stable estimates for the cycle.

Next, in Table 10 we present the values of the model for the potential infections at the specific days of the cycle (per 43 days). These days must be evaluated the situation: performance of strategies for controlling the outbreak.

A crude estimate is that the minimum of outbreak (at the start of each cycle) is reported on day 130 (start of cycle 4). If the preventing strategies remain the same and the external factors as weather are assumed as not significant for the evolution of the situation, this point can be a starting estimation for the termination of the outbreak (or at least for its downward trend).

A rational question should be that why do not use only tSIR model for modeling an outbreak instead of using it only as part of another approach. In order to display the superiority of forecasts of the proposed approach compared to tSIR epidemiological model, Table 11 displays the head to head comparison of forecasts from the tSIR model and the proposed approach according to MAPE criterion for the in-sample last 7 observations.

It is obvious that the tSIR model cannot compete in terms of forecasting accuracy the suggested approach. However, is particularly useful for extracting some metrics, i.e. estimation of downtrend of the outbreak and expected reported rate. These metrics are suitable for comparison of the situation between countries.

4.4. Comparison of Greece with other countries

After the application of the suggested method to Greece, the aim is to compare Greece with other countries and the whole World based on the expected reported rate (cumulative cases over cumulative births estimation from the model), the estimation of the downward trend for the outbreak and the basic reproduction number. The comparison is based on the tSIR model using the last reported Population and birth rates (crude per 1000 persons) data from World Bank. Moreover, we estimate the reproducible number of the disease (basic reproduction number) using the Exponential Growth method with R package R0. Table 12 displays the results.

We consider as key points for the comparison of the situation of countries: the expected reported rate, the estimation of the outbreak reversion and the basic reproduction number. in order to be decided a single metric which rank the countries from these which seem to have the slighter problem, thus these which can be considered as more effective (lowest value), to the countries with the greater problematic...
situation and these countries can be considered as less effective (highest value), we use the following procedure:

1. **Rank the countries according to each variable**

2. **i) Expected Reported Rate, ii) Estimation of Outbreak Reversion, iii) Basic Reproduction Number**

3. **Take the average Position (simple mean) of the rankings**

The ranking of the countries according to each variable and overall is displayed in **Table 13** and in **Fig. 3**.

The expected reported rate (i.e. cumulative cases over cumulative births) is almost 300 in the USA, over 150 in Sweden, almost 70 in Spain, over 55 in Italy, over 40 in the UK, between 20 and 30 for Switzerland and Netherlands and below 20 for the rest of the countries. The basic reproduction number is lower than 1 for all countries except for France and USA. The expected reported rate is highest in USA and lowest in Finland.

### Table 9
**Expected Reporting Rate and Estimated Cycle Length.**

| Country | Expected Reporting Rate | Average Cycle Length | Basic reproduction Number ($R_0$) |
|---------|-------------------------|----------------------|-----------------------------------|
| Greece  | 11.332                  | 43                   | 1.828                             |

### Table 10
**Number of Expected Infected Persons (start of cycle/day).**

| Country  | Expected Reporting Rate | Cycle Length | Basic Reproduction Number |
|----------|-------------------------|--------------|---------------------------|
| Finland  | 4.260 (44)              | 3.923 (216)  | 1.828                     |
| Greece   | 3.738 (87)              | 9.008 (259)  |                           |
| Switzerland | 2.387 (130)        | 3.920 (302)  |                           |
| Finland  | 3.141 (173)             | 3.339 (345)  |                           |

### Table 11
**Forecasting Comparison Between tSIR model and Suggested Approach.**

| Criterion | Suggested Approach | tSIR Model |
|-----------|--------------------|------------|
| MAPE      | 0.0026             | 0.1199     |

The ranking of the countries according to each variable and overall is displayed in **Table 13** and in **Fig. 3**.

### Table 13
**Relative Effectiveness of Countries.**

| Country | Expected Reported Rate | Estimation of Outbreak Reversion | Basic Reproduction Number | Average Position |
|---------|------------------------|----------------------------------|---------------------------|------------------|
| Finland | 4                      | 1                                | 3                         | 2,667            |
| Germany | 5                      | 2                                | 5                         | 4,000            |
| Turkey  | 3                      | 3                                | 7                         | 4,333            |
| Austria | 10                     | 5                                | 1                         | 5,333            |
| Greece  | 2                      | 11                               | 4                         | 5,667            |
| Netherlands | 7                      | 9                                | 6                         | 7,333            |
| France  | 6                      | 7                                | 10                        | 7,667            |
| UK      | 9                      | 8                                | 8                         | 8,333            |
| Spain   | 11                     | 6                                | 9                         | 8,667            |
| Sweden  | 12                     | 5                                | 12                        | 9,667            |
| USA     | 13                     | 5                                | 13                        | 10,333           |

### Table 12
**Comparison of Expected Reporting Rates and Estimated Cycle Lengths.**

| Country | Expected Reported Rate | Fitting measure $R^2$ | Average Cycle Length | Estimation of Outbreak Reversion | Basic Reproduction Number |
|---------|------------------------|-----------------------|----------------------|----------------------------------|---------------------------|
| Greece  | 11.332                 | 0.47                  | 43                   | 130                              | 1.828                     |
| Italy   | 56.619                 | -0.01                 | 53                   | 54                               | 0.145                     |
| Spain   | 68.031                 | 0.50                  | 27                   | 55                               | 0.701                     |
| Germany | 19.505                 | 0.07                  | 8                    | 41                               | 0.385                     |
| Switzerland | 27.721           | 0.64                  | 26                   | 53                               | 0.188                     |
| Austria | 16.502                 | 0.52                  | 50                   | 351                              | 0.344                     |
| Finland | 17.461                 | 0.26                  | 1                    | 39                               | 0.246                     |
| Sweden  | 153.197                | 0.42                  | 53                   | 54                               | 2.236                     |
| Turkey  | 17.006                 | 0.50                  | 50                   | 51                               | 0.509                     |
| Netherlands | 24.736             | 0.01                  | 91                   | 92                               | 0.458                     |
| UK      | 40.997                 | 0.02                  | 6                    | 73                               | 0.589                     |
| USA     | 298.954                | 0.55                  | 53                   | 54                               | 4.182                     |
| France  | 22.685                 | 0.51                  | 7                    | 64                               | 0.824                     |
| World   | 31.634                 | 0.39                  | 50                   | 51                               | 7.203                     |

C. Katris
reproduction number displays the number of affected people by one single patient. It appears to be over 4 in the USA, over 2 in Sweden, over 1.8 in Greece and below 1 for the rest of the countries. The estimation of the reversion which could mean termination or significant reduction of the disease and Austria, Greece and Netherlands are the most pessimistic forecasts. Greece while at an early stage displayed smaller values of these metrics, thus smaller dispersion of the virus compared to other countries, now with a Basic Reproduction number almost 2 and estimation of outbreak reversion after 130 days lies in the middle of the ranking. Spain and Italy were the countries with severe results of the disease and Austria, Greece and Netherlands are the most pessimistic forecasts. Greece while at an early stage displayed smaller values of these metrics, thus smaller dispersion of the virus compared to other countries, now with a Basic Reproduction number almost 2 and estimation of outbreak reversion after 130 days lies in the middle of the ranking. Spain and Italy were the countries with severe results of the disease and Austria, Greece and Netherlands are the most pessimistic forecasts. Greece while at an early stage displayed smaller values of these metrics, thus smaller dispersion of the virus compared to other countries, now with a Basic Reproduction number almost 2 and estimation of outbreak reversion after 130 days lies in the middle of the ranking.

Table 14 displays the name of the country in the first column, the value of the Stringency index, the ranking of the Stringency index and the ranks of Expected Reported Rate, Estimation of the reversion of the outbreak, of basic reproduction number and finally the average position of these metrics which represent the overall efficiency of each country.

With the aim to decide whether the strictness of measures had an impact, is computed the correlation between the ranks of Stringency Index with the ranks of expected reported rate, of estimation of outbreak reversion, of basic reproduction number and of average position which represents the overall effectiveness of the country. The ranks of the values are compared using Spearman Correlation. Table 15 displays the results (in parenthesis are reported the p-values of the null hypothesis that there is no correlation between the variables).

From Table 15, is observed that the correlations between the variables are weak and in case of estimation of outbreak reversion almost zero. The corresponding p-values display that there is not significant

Table 15
Correlations Between Stringency Index and Other Variables.

|                      | Expected Reported Rate | Estimation of Outbreak Reversion | Basic Reproduction Number | Effectiveness - Avg. Position |
|----------------------|------------------------|---------------------------------|---------------------------|-------------------------------|
|                      | 0.3352 (0.263)         | 0.0166 (0.957)                  | 0.3407 (0.255)            | 0.3246 (0.279)                |

Tracking of government responses related to lockdown exiting, one can visit the following sites:
1) https://ourworldindata.org/policy-responses-covid
2) https://ourworldindata.org/grapher/covid-stringency-index
3) https://ig.ft.com/coronavirus-lockdowns/

An analytical view and also comparisons of policies to the fighting against the disease, stringency index for government response and for a

Fig.3. Relative Effectiveness of Countries.

4.4.1. Presence of additional strictness – e.g. A lockdown

In this section, is attempted to find whether the strictness of a country -including the presence of a lockdown- is affected the performance of the country against the pandemic. The response of each country to the threat by employing a lockdown is different. In order to compare the effectiveness of one country’s policies over time or the impact of policies between countries, there are used many different measures and is clearly difficult to be achieved such type of comparison.

A team at Oxford university’s Blavatnik School of Government is maintaining a database of pandemic-response policies and using it to derive an index of the measures’ overall stringency, with the aim to allow such comparisons. This index is named Government Response Stringency Index. An analysis of the variation in government responses to COVID-19 takes place on (Hale et al, 2020).

An analytical view and also comparisons of policies to the fighting against the disease, stringency index for government response and for a
correlation. This means that we cannot conclude that additional strictness is correlated with the effectiveness of the countries in fighting the outbreak. However, the sample is small and additional research is needed.

5. Summary and conclusions

In this paper we made an attempt to build a time-series based statistical data-driven method for outbreak monitoring with the use of time series models, probability distributions and an epidemiological model. After the detailed description of the method, it follows an application to the COVID-19 outbreak for the case of Greece. We generate forecasts, alternative scenarios and provide an estimate for the termination or at least the beginning of the downward trend for the outbreak. For the application at the case of Greece, forecasts are made with the Newbolt/Granger combination scheme of TES, ARIMA (with normal innovations) and ANN (with 1 input node) models, scenarios are built with the help of Log-normal distribution, estimates for the expected reported rate and for the downward trend of the outbreak are made through the tSIR epidemiological model and is calculated the basic reproduction factor by using the Exponential Growth method. Finally, is applied a framework in order to compare the severity of the situation between different countries and is discussed the presence of additional strictness - e.g. a lockdown.

CRediT authorship contribution statement

Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing - original draft, Writing - review & editing.

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.

References

Ahmadi, A., Fadai, Y., Shirani, M., & Rahmani, F. (2020). Modeling and Forecasting Trend of COVID-19 Epidemic in Iran. medRxiv.

Ahmed, M. A., Hassan, M. A., & Al-El Aziz, M. (2020). Optimization method for forecasting confirmed cases of COVID-19 in China. Journal of Medical Case Reports, 9 (3), 674.

Anastassopoulou, C., Russo, L., Tsakris, A., & Siettos, C. (2020). Data-based analysis, modelling and forecasting of the COVID-19 outbreak. PeerJ One, 15(3), Article e0230405.

Anderson, R. M., & May, R. M. (1991). Infectious diseases of humans: dynamics and control. Oxford university press.

Becker, A. D., & Grenfell, B. T. (2017). ts:R: An R package for time-series Susceptible-Infected-Recovered models of epidemics. PeerJ One. 12(9).

Boelle, P. Y., Obadia, T., & Obadia, M. T. (2015). Package ‘R0’.

Box, G. E., & Jenkins, G. M. (1976). Time series analysis: Forecasting and control. San Francisco: Holden Day.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (1994). Time series analysis: Forecasting and control. John Wiley & Sons.

Caudron, Q., Mahmud, A. S., Metcalf, C. J. E., Goffe, N., Viboud, C., Cliff, A. D., & Grenfell, B. T. (2015). Predictability in a highly stochastic system: Final size of measles epidemics in small populations. Journal of The Royal Society Interface, 12 (102), 20141125.

Delignette-Muller, M. L., Dutang, C., Pouillot, R., Denis, J. B., & Silberchicot, A. (2019). Package ‘fitdistrplus’.

Earnest, A., Chen, M. J., Ng, D., & Sin, L. Y. (2005). Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore. BMC Health Services Research, 5(1), 36.

Fanelli, D., & Piazza, F. (2020). Analysis and forecast of COVID-19 spreading in China, Italy and France. Chaos, Solitons & Fractals, 134, Article 109761.

Finkenstadt, B. F., & Grenfell, B. T. (2000). Time series modelling of childhood diseases: A dynamical systems approach. Journal of the Royal Statistical Society. Series C (Applied Statistics), 49(2), 187–205.

Friedman, J. H. (1991). Multivariate adaptive regression splines. The Annals of Statistics, 1–67.

Hale, T., Petherick, A., Phillips, T., & Webster, S. (2020). Variation in government responses to COVID-19. Blavatnik School of Government, Working Paper, 31.

Haykin, S. (2010). Neural networks: a comprehensive foundation. 1999. Mc Millan, New Jersey, 1-24.

Holt, C. C. (2004). Forecasting seasonalities and trends by exponentially weighted moving averages. International Journal of Forecasting, 20(1), 5–10.

Hu, Z., Qiyang, G., Shudi, L., Li, J., & Momiao, X. (2020). Artificial Intelligence Forecasting of Covid-19 in China. arXiv preprint, arXiv:2002.07112.

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice. OTexts, Hyndman, R. J., & Khandakar, Y. (2007). Automatic time series for forecasting: the forecast package for R (No. 6/07). Clayton VIC, Australia: Monash University, Department of Econometrics and Business Statistics.

Kim, M. J., Price, N., Scotch, M., & Rabinowitz, P. (2014). Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. BMC Bioinformatics, 15(1), 276.

Lee, T. H., White, H., & Granger, C. W. (1993). Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests. Journal of Econometrics, 56(3), 269–290.

Lewis, P. A., & Stevens, J. G. (1991). Nonlinear modeling of time series using multivariate adaptive regression splines (MARS). Journal of the American Statistical Association, 86(416), 864–877.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., … Xing, X. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. New England Journal of Medicine.

M.C. Limar J.B.O. Mere A.G. Marcos F.D.P. Ascarci A MORE flexible neural network package (Package ‘AMORE’) 2014.

Lippmann, R. (1987). An introduction to computing with neural nets. IEEE Assp Magazine, 4(2), 4–22.

Meyer, A., Kenny, G., & Quinn, T. (1998). Forecasting Irish inflation using ARIMA models.

Milbourne, M. S. (2019). Package ‘earth’. R Software package.

Petropoulos, F., & Makridakis, S. (2020). Forecasting the novel coronavirus COVID-19. PeerJ One, 15(3), Article e021236.

Prompong, S., Jaroonutsatinee, M., & Jaroonutsatinee, K. (2006). Forecasting Dengue Haemorrhagic Fever Cases in Southern Thailand using ARIMA Models.

Roostaei, K., Lee, Y., Luo, R., Karpich, A., Rothenberg, R., Hyman, J. M., … Chowell, G. (2020). Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. Infectious Disease Modelling, 5, 256–263.

Takahashi, S., Liao, Q., Van Boeckel, T. P., Xing, W., Sun, J., Hsiao, V. Y., … Wu, J. T. (2016). Hand, foot, and mouth disease in China: Modeling epidemic dynamics of enterovirus serotypes and implications for vaccination. PLoS Medicine, 13(2).

Tetko, I. V., Livingstone, D. J., & Luik, A. I. (1995). Neural network studies. 1. Comparison of different methods. Computer Methods and Programs in Biomedicine, 45(3), 187–199.

Tetko, I. V., Livingstone, D. J., & Luik, A. I. (1995). Neural network studies. 2. Comparison of different methods. Computer Methods and Programs in Biomedicine, 45(3), 269–283.

Van den Driessche, P., & Watmough, J. (2008). Further notes on the basic reproduction number. In Mathematical epidemiology (pp. 159–178). Berlin, Heidelberg: Springer.

Wallings, J., & Lipshitz, M. (2007). How generation intervals shape the relationship between growth rates and reproductive numbers. Proceedings of the Royal Society B: Biological Sciences, 274(1699), 599–604.

Weiss, C. E., Roetter, G. R., Raviv, E., & Weiss, M. C. E. (2018). Package ‘ForecastComb’.

Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential national and international spread of novel 2019-nCoV outbreak originating in Wuhan, China: A modelling study. The Lancet, 395(10225), 689–697.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. International Journal of Forecasting, 14(1), 35–62.

Zhao, S., Liu, Q., Fan, J., Masa, S. S., Yang, G., Wang, W., … Wang, M. H. (2020). Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. International Journal of Infectious Diseases, 92, 214–217.