

Article

Predicting the Spread of SARS-CoV-2 in Italian Regions: The Calabria Case Study, February 2020–March 2022

Francesco Branda ^{1,†} , Ludovico Abenavoli ^{2,*,†} , Massimo Pierini ^{3,4,†}  and Sandra Mazzoli ^{4,†} 

¹ Department of Computer Science, Modeling, Electronics and Systems Engineering (DIMES), University of Calabria, 87036 Rende, Italy; francesco.branda@unical.it

² Department of Health Sciences, University Magna Graecia, 88100 Catanzaro, Italy

³ Guglielmo Marconi University, 00193 Rome, Italy; info@epidata.it

⁴ SITO WEB del Gruppo Epidemiologico, EpiData.it, 24121 Bergamo, Italy; sandra.mazzoli50@gmail.com

* Correspondence: l.abenavoli@unicz.it

† These authors contributed equally to this work.

Abstract: Despite the stunning speed with which highly effective and safe vaccines have been developed, the emergence of new variants of SARS-CoV-2 causes high rates of (re)infection, a major impact on health care services, and a slowdown to the socio-economic system. For COVID-19, accurate and timely forecasts are therefore essential to provide the opportunity to rapidly identify risk areas affected by the pandemic, reallocate the use of health resources, design countermeasures, and increase public awareness. This paper presents the design and implementation of an approach based on autoregressive models to reliably forecast the spread of COVID-19 in Italian regions. Starting from the database of the Italian Civil Protection Department (DPC), the experimental evaluation was performed on real-world data collected from February 2020 to March 2022, focusing on Calabria, a region of Southern Italy. This evaluation shows that the proposed approach achieves a good predictive power for out-of-sample predictions within one week (R-squared > 0.9 at 1 day, R-squared > 0.7 at 7 days), although it decreases with increasing forecasted days (R-squared > 0.5 at 14 days).



Citation: Branda, F.; Abenavoli, L.; Pierini, M.; Mazzoli, S. Predicting the Spread of SARS-CoV-2 in Italian Regions: The Calabria Case Study, February 2020–March 2022. *Diseases* **2022**, *10*, 38. <https://doi.org/10.3390/diseases10030038>

Academic Editor: Amal K. Mitra

Received: 30 May 2022

Accepted: 28 June 2022

Published: 30 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: SARIMA; time series regression models; forecasting; epidemiology; COVID-19; SARS-CoV-2; Italy; Calabria

1. Introduction

In December 2019, the new Coronavirus, SARS-CoV-2 [1], emerged and into a world population without proper specific immunization. Due to its high infectivity, the virus spread worldwide, beginning the new current pandemic [2]. During the epidemic and subsequent vaccine immunization, one of the main issues was the emergence of several Variants of Concern (VOC) [3] with consequent re-emergence of new infection cases. In particular, several notable variants of SARS-CoV-2 have emerged in recent months [4], such as B.A.2 [5], which expresses a very high infectivity rate and pathogenesis, and B.A.2.2 [6], responsible in Hong Kong for the significant increase in lethality among the unvaccinated and uninfected elderly and children.

This highlighted the importance of following the contagion evolution on a temporal scale inside the various populations, especially to define the restarting moment of new cases to prevent severe clinical cases, along with ICU admissions and deaths, to determine needed non-pharmacological intervention (NPI), to test the success or failure of containment measures in place, and to guide governments and decision makers. Moreover, it has become increasingly important to define and refine epidemiological forecasting methods in order to adequately monitor population infection cases and deaths. SARIMA(X) models are an example of methods that have already been used for prediction in epidemiology, such as malaria [7], influenza-like illness [8], dengue hemorrhagic fever [9], West Nile virus [10], scarlet fever [11], human brucellosis [12], and recently, COVID-19 [13,14].

This paper presents the design and implementation of an approach based on autoregressive models to reliably forecast the spread of COVID-19 in Italian regions. The method automatically collects daily data on the number of individuals infected with SARS-CoV-2 and performs the following operations: (i) *data pre-processing*, which consists of turning raw data into the required format for the purposes of analysis; (ii) *predictive modeling*, which regards the training of a model to forecast the number of infections that will happen in a specific area; (iii) *results visualization*, which presents the results in a graphical way, allowing users to visually explore the data.

As a case study, we present here the analysis of the regional trend of SARS-CoV-2 spread in Calabria, a region of Southern Italy, based on open-access coronavirus data provided by the Italian Civil Protection Department (DPC), starting from 24 February 2020. The results of the experimental evaluation show the effectiveness of the method, by achieving good accuracy in new infections forecasting within one week.

The rest of the paper is organized as follows. Section 2 outlines the problem statement and describes the proposed approach in detail. Section 3 presents the experimental evaluation. Finally, Sections 4 and 5 conclude the paper and plan future research works.

2. Materials and Methods

We begin by outlining a proper notation to be used throughout the paper. Let D be a dataset collecting epidemiological data, where each d_i is described by the following tuple: $\langle ID, T, NC, DE, RC, HOSP, ICU, TESTS \rangle$, where ID is the identifier of a region, T is the notification date of an infection, NC indicates new positive cases, DE is the total amount of deaths, RC refers to the total number of recovered persons, $HOSP$ and ICU are hospitalized patients with symptoms and in intensive care, respectively, and $TESTS$ is the number of tests performed. Let $\hat{T} = \langle \hat{t}_1, \hat{t}_2, \dots, \hat{t}_K \rangle$ be an ordered timestamp list and $H = \langle \hat{t}_j, \hat{t}_{j+1}, \dots \rangle$ a future temporal horizon, with $j > K$.

As mentioned before, the main goal of this work is to reliably predict the number of new infections at a given timestamp $\hat{t}_j \in K$. This is achieved in two steps: (i) estimating the COVID-19 epidemic risk in Italian regions for identifying high-risk areas (i.e., areas with particularly high risk of infection due to a high incidence of spread of SARS-CoV-2); (ii) training a model that given a timestamp $\hat{t}_j \in H$ states the number of new cases $N \in NC$ that are predicted to happen in such areas at the timestamp \hat{t}_j .

Figure 1 presents the general idea of the approach through a graphic representation of the whole process as a sequence of three steps. The input data of the analysis is the set of collected epidemiological data to be processed. The second step consists of cleaning, selecting, and transforming raw data into the desired format so that useful information can be derived from it. In other words, (i) incorrect and incomplete data are removed; (ii) a subset of data is selected to make it suitable for analysis; (iii) third-party data from an external authoritative source are merged with the existing database to enrich collected data. The third step is aimed at detecting the regions mostly affected by the pandemic and extracting a prediction model for providing dynamic monitoring of the spread of the disease, and supporting organizations in the evaluation of the effect of local containment measures. Finally, results visualization is performed by using an interactive dashboard to inform citizens in an intuitive way and make the collected data available.

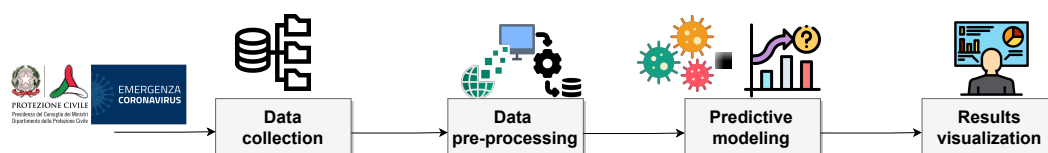


Figure 1. Proposed approach steps.

Figure 2 reports the meta-code of the *predictive modeling* step. In particular, given a specific epidemiological dataset, the *EstimateTransmission()* method estimates the net reproduction numbers (R_t) (i.e., the average number of new cases generated by an infectious

case at a given time of the epidemic) based on confirmed cases reported to the National Integrated Surveillance System (<https://www.epicentro.iss.it/en/coronavirus/sars-cov-2-integrated-surveillance>, accessed on 28 May 2022) and stratified by region. Methodological details can be found in [15].

ALGORITHM

Input :Epidemiological dataset D

Output: Net Reproduction Numbers R_t Predictive Model M

```

1  $\{RR_{t_1}, \dots, RR_{t_K}\} \leftarrow EstimateTransmission(D)$ 
2  $\{\hat{D}_1, \dots, \hat{D}_K\} \leftarrow BuildTSData(D, RR_{t_1}, \dots, RR_{t_K})$ 
3 foreach  $k = 1, \dots, K$  do
4    $M_i \leftarrow GeneratePredictionModel(\hat{D}_i)$ 
5    $M \leftarrow M \cup M_i$ 
6 end
7 return  $(RR_t, M)$ 

```

Figure 2. Workflow of the *predictive modeling* step.

As soon as this step is completed, the epidemiological dataset D is transformed in K time series datasets. Specifically, this task is executed by the *BuildTSData()*, which transforms D in the time series dataset collection $\hat{D} = \{\hat{D}_1, \dots, \hat{D}_K\}$ where each \hat{D}_i is the time series of new cases of COVID-19 in a region $RR_{ti} \in RR_t$.

Finally, for each \hat{D}_i , the *GeneratePredictionModel()* method generates a model M to forecast the number of cases that will happen in the specific region RR_{ti} . This task is performed using *SARIMA*(p, d, q)(P, D, Q)[s] model, i.e., a further development of Autoregressive Integrated Moving Average *ARIMA*(p, d, q) with seasonality, where p is the number of AR autoregression terms, d is the difference order, q is the number of MA sliding average terms, P refers to the maximum lag order of the seasonal autoregression term, Q is the maximum lag order of the moving average operator, D is the seasonal difference order, and s is the seasonal difference cycle step [16]. *SARIMA*(p, d, q)(P, D, Q)[s] corresponds to the polynomial operators formula

$$\phi_p(B)\Phi_P(B^s)(1 - B)^d(1 - B^s)^D x_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \tag{1}$$

where B is the backshift operator (or lag operator) and $B^n x_t = x_{t-n}$; $\phi_p(B)$ and $\Phi_P(B^s)$ are the polynomial operators of AR autoregression (the latter refers to seasonality), $\theta_q(B)$ and $\Theta_Q(B^s)$ are the polynomial operators of MA sliding average (the latter refers to seasonality), where a generic polynomial operator $\Omega_m(B^n)$ yields

$$\Omega_m(B^n) = 1 - \sum_{i=1}^m \Omega_i B^{in} \tag{2}$$

The best model, balancing between complexity and adjustment, has been chosen via a grid-search to optimize the Bayesian Information Criterion (BIC) [16] resulting in a *SARIMA*(0, 1, 1)(1, 0, 1)(7) corresponding to the polynomial operators formula

$$\phi_0(B)\Phi_1(B^7)(1 - B)^1(1 - B^7)^0 x_t = \theta_1(B)\Theta_1(B^1)\varepsilon_t \tag{3}$$

$$\Phi_1(B^7)(1 - B)x_t = \theta_1(B)\Theta_1(B^7)\varepsilon_t \tag{4}$$

$$(1 - \Phi_1 B^7)(1 - B)x_t = (1 - \theta_1 B)(1 - \Theta_1 B^7)\varepsilon_t \tag{5}$$

$$(1 - B - \Phi_1 B^7 + \Phi_1 B^8)x_t = (1 - \Theta_1 B^7 - \theta_1 B + \Theta_1 \theta_1 B^8)\varepsilon_t \tag{6}$$

$$x_t - x_{t-1} - \Phi_1 x_{t-7} + \Phi_1 x_{t-8} = \varepsilon_t - \Theta_1 \varepsilon_{t-7} - \theta_1 \varepsilon_{t-1} + \Theta_1 \theta_1 \varepsilon_{t-8} \tag{7}$$

which yields the explicit formula

$$x_t = x_{t-1} + \Phi_1(x_{t-7} - x_{t-8}) + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \Theta_1(\varepsilon_{t-7} - \varepsilon_{t-8}) \tag{8}$$

where x_t is the observed variable at time t , ε_t is the white noise, identically and independently normally distributed with mean 0 and variance σ^2 , and Φ_1 , θ_1 , and Θ_1 are the parameters to be estimated. The chosen model shows fairly acceptable diagnostics (standardized residuals, estimated density, Q–Q plot, and correlogram) [17,18], as can be seen in Figure 3.

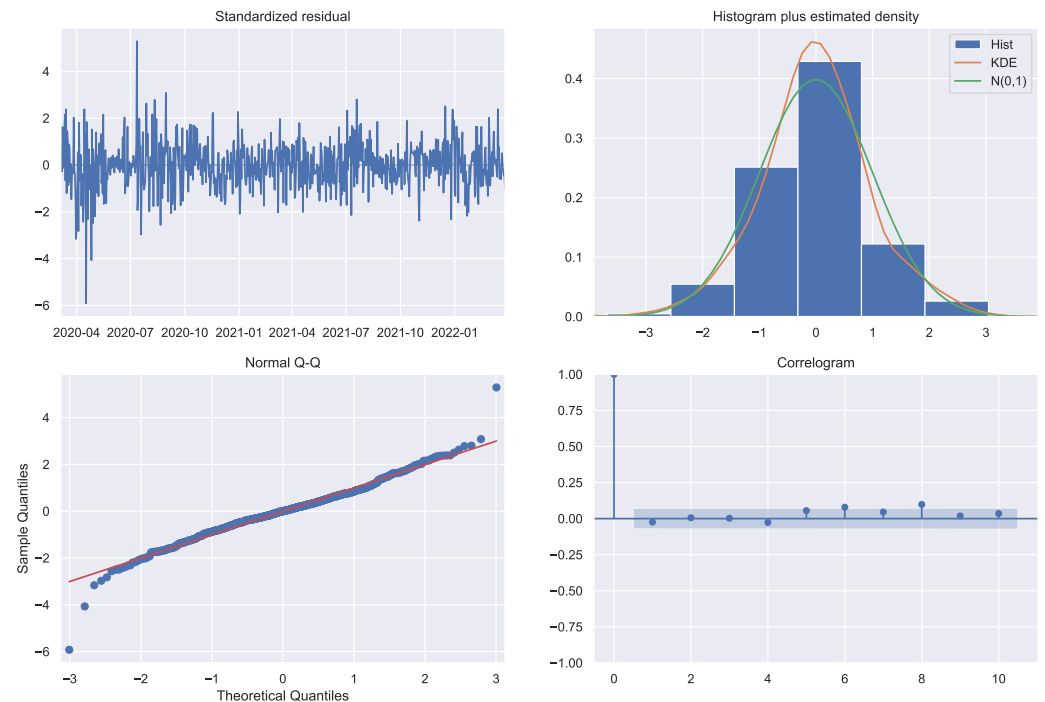


Figure 3. SARIMA model diagnostics.

Software

All analyses were performed in Python 3.8.3 with: *statsmodels* (<https://www.statsmodels.org/stable/index.html>, accessed on 28 May 2022), *pmdarima* (<https://pypi.org/project/pmdarima/>, accessed on 28 May 2022) and *scipy* (<https://scipy.org/>, accessed on 28 May 2022). Plots have been created with *matplotlib* (<https://matplotlib.org/>, accessed on 28 May 2022) and *seaborn* (<https://seaborn.pydata.org/>, accessed on 28 May 2022). The complete Jupyter notebook Python code is publicly available on GitHub at https://github.com/maxdevblock/covid19_sarima_calabria, accessed on 28 May 2022.

3. Results

The data that we used to evaluate the effectiveness and accuracy of the approach described above are gathered from the Italian Civil Protection Department (DPC) [19] dataset and enriched at the provincial level by using an automatic scraper that extracts and transforms data from the website of the Calabria region (<https://regione.calabria.it/website/>, accessed on 28 May 2022) into a machine-readable format in order to make its reuse easier [20] (text in Italian). The database is freely accessible at <https://github.com/fbranda/covid19-opensdata-calabria>, accessed on 28 May 2022. Moreover, data are graphically consultable on the COVIDA platform at <https://covidml.com/>, accessed on 28 May 2022.

We carried out an experimental evaluation by analyzing the SARS-CoV-2 trend in the period between 24 February 2020 and 27 March 2022, in Calabria, one of the Italian regions, which has been most affected during the fourth wave of the pandemic.

Figure 4 shows a preliminary view of the collected epidemiological data. Specifically, Figure 4A reports the time plot of the new confirmed cases, in which new positive cases are plotted versus the time of notification. From the plot, we see that the number of cases

exhibits a stable trend until December 2021, followed by a sharp increase in infections in January 2022, due to the circulation of the SARS-CoV-2 Omicron variant.

Omicron's higher infection rate has pushed health systems to the breaking point, causing a significant number of deaths and hospitalizations. A clearer view of the Omicron's wave can be seen in Figure 4B,C, which show the number of confirmed COVID-19 deaths and hospitalizations per day, respectively. In particular, the new record for deaths in one day was 18 on 28 January 2022 (previously it was 17 on 23 November 2020), whereas the hospitalizations have risen significantly, approaching the values of the prior wave (446 hospitalizations on 18 January 2022 versus 482 hospitalizations on 26 April 2021).

However, for the innate features of Omicron and the protection offered by the vaccines, a smaller percentage of COVID patients were admitted to intensive care units (ICUs), as shown in Figure 4D. The chart clearly shows that during the first phase of Omicron surge, an initial increase in patients (with 38 persons admitted to ICU on 11 January 2022) was followed by a smooth decreasing trend.

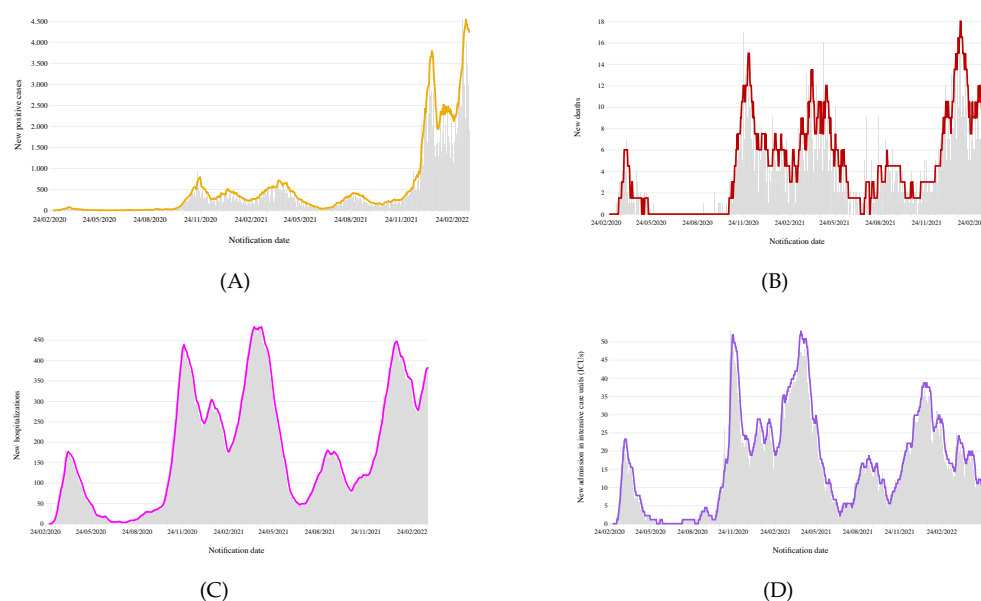


Figure 4. Calabria epidemiological data: (A) new positive cases; (B) total amount of deaths; (C) hospitalized patients with symptoms and (D) in intensive care.

To evaluate the impact of SARS-CoV-2 circulation, we analyzed the net reproduction number R_t (i.e., the transmission potential at a given time t of the epidemic once interventions are introduced or the susceptibility in the population decreases), calculated by the research group CovidStat INFN (<https://covid19.infn.it/progetto.html>, accessed on 28 May 2022). This method uses the growth rate determined over the last 14 days with an exponential fit to the number of infected persons per day [21] assuming the mean value of the generation time published by Cereda et al. [22]. As shown in Figure 5, with the national lockdown imposed in early March 2020, R_t estimates followed a constantly decreasing trend. Since late May 2020 with the gradual reopening of all activities, R_t started to fluctuate, reaching maximum values around 3 in the week from 3 to 10 August. From 7 January 2021 to 27 March 2022, R_t remained nearly constant at values around 1.5–1.8.

To forecast new COVID-19 cases for the next 14 days, a seasonal autoregressive integrated moving average without exogenous variables (SARIMA) model was used. The data are characterized by heteroskedasticity (Breusch–Pagan $p \ll 0.01$), non-stationarity (Dickey–Fuller $p = 0.76$), and seasonality (7 day cycles), as shown in Figure 6.

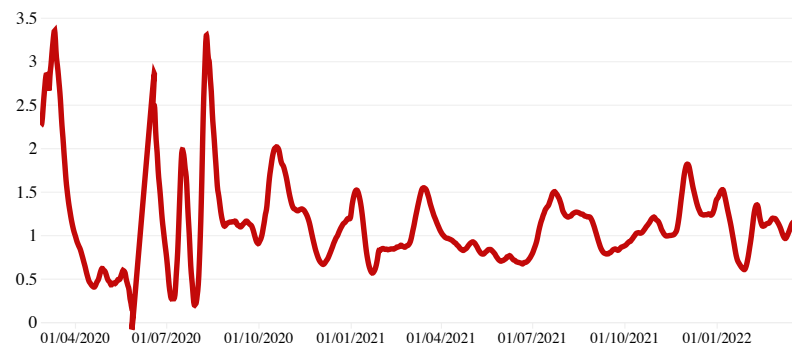


Figure 5. COVID-19 estimated R_t in Calabria over a 7-day moving average, 24 February 2020–27 March 2022.

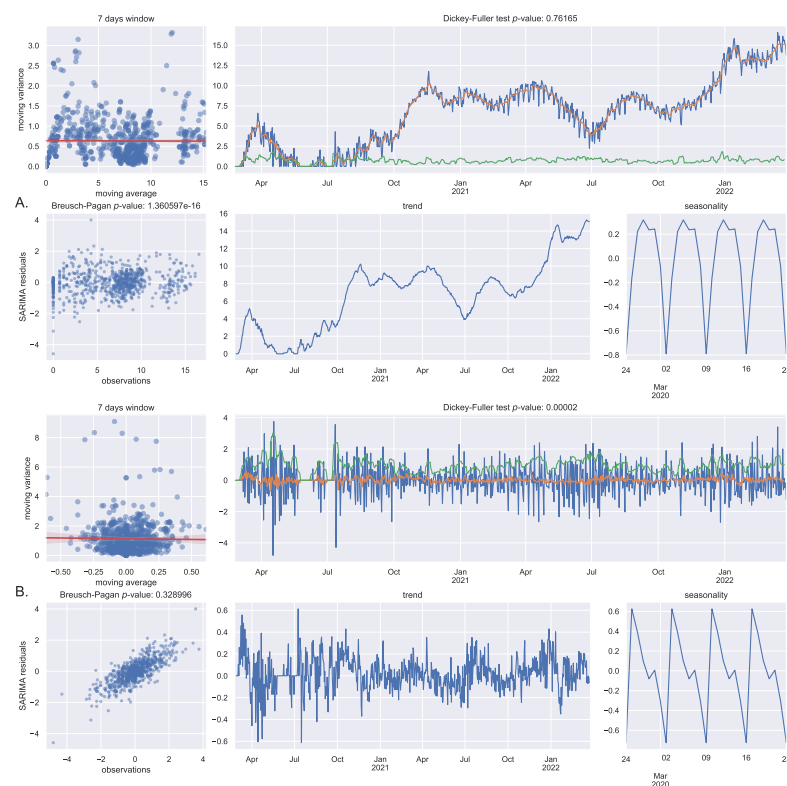


Figure 6. Data tests: (A) Box–Cox transformed data; (B) Box–Cox transformed and differentiated data.

We call “seasonality” the 7-days (circaseptan) observed oscillations because it has been treated as a seasonal behavior in the SARIMA model, even if, usually in epidemiology, the seasonality refers to longer periods (such as months). In some countries, additional hemicaseptan (3.5 days) and 14-days periodicities have been observed [23]. The circaseptan seasonality of COVID-19 new daily cases, which has not been observed or reported in prior epidemics, is believed to be likely associated with epidemiological and social factors, mainly testing bias and reporting bias [23–25].

Since, as stated above, such a circaseptan seasonality has not been observed previously in prior epidemics, this study cannot be compared to prior studies but similar SARIMA models have been recently tested using circaseptan seasonality to forecast 14 or 28 days [13,14,26] with comparable results.

We have chosen not to solve heteroskedasticity but to treat it as an inherent feature also because a transformation can damage results interpretability and not completely solve the issue [27] that could be reduced with differentiation only. Nevertheless, the data have been transformed with the Box–Cox method in order to avoid negative results [28].

Non-stationarity can be solved by first-order differentiation, i.e., SARIMA parameter $d = 1$ (Dickey–Fuller $p \ll .01$). As expected, differentiating can partially solve heteroskedasticity too (Breusch–Pagan $p = 0.33$). The 7-days cycle seasonality has been confirmed with both seasonal decomposition and periodogram [29] that clearly shows first and second harmonic (7 and 3.5 days). Confidence intervals of 50% and 90% have been chosen for out-of-sample 14 days’ new cases predictions (see Figure 7).

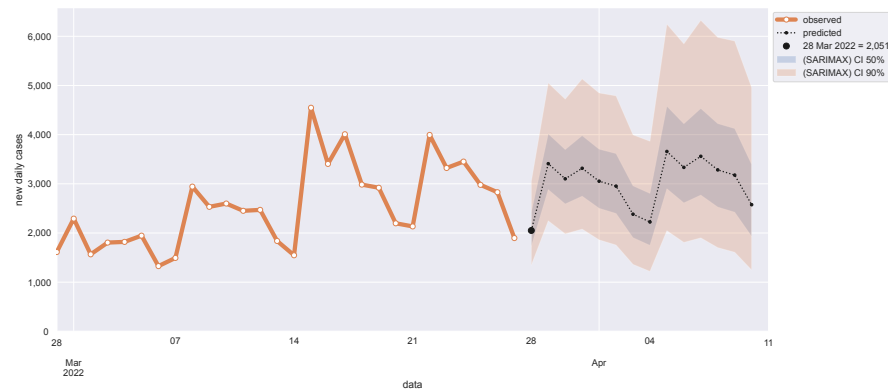


Figure 7. Out-of-sample 14-days prediction of daily new COVID-19 cases in Calabria with SARIMA model.

To define the SARIMA model R-squared (R^2) score for out-of-sample 14-days prediction, for each day between 1 January 2021, and 13 March 2022, we have chosen the best SARIMA model via grid-search (based on previous observations) and forecasted the next 14 days. Pooled R^2 score of 14 days of observations and predictions is 0.90 at 1 day, greater than 0.80 within 6 days, and greater than 0.50 up to 14 days (see Figure 8 and Table 1). These results confirm the appropriateness of the autoregressive model and its good performance in the epidemiology domain over rolling time horizons.

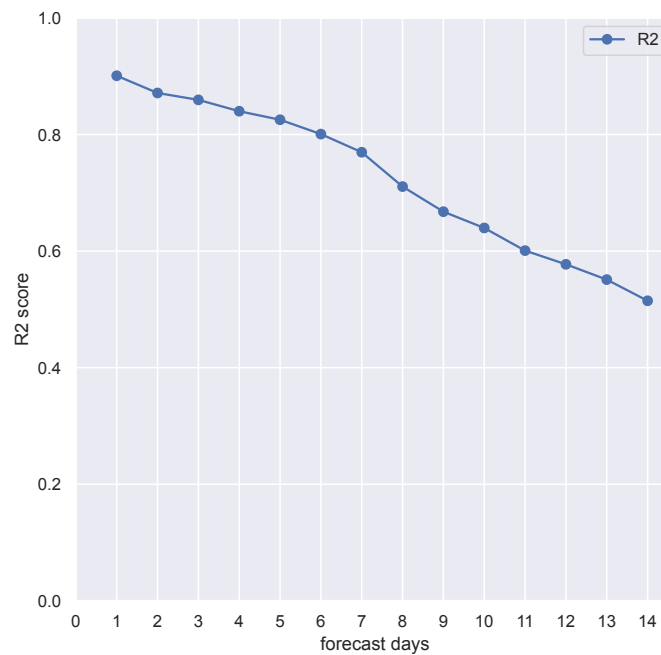


Figure 8. Pooled R^2 scores of the 14-days out-of-sample forecast of COVID-19 new daily cases in Calabria (Italy) with SARIMA model.

Table 1. Pooled R^2 score for each forecasted day.

Forecasted Days	R^2
1	0.90
2	0.87
3	0.86
4	0.84
5	0.82
6	0.80
7	0.77
8	0.71
9	0.67
10	0.64
11	0.60
12	0.58
13	0.55
14	0.51

4. Discussion

Italy was the second country to have a large outbreak of infections of novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), with clusters of cases detected in Lombardy and Veneto on 21 February 2020, and the first deaths on 22 February 2020 [30,31]. By the beginning of March, the virus had spread to all regions of Italy, and to reduce the burden of the epidemic on the healthcare settings, the government imposed a national lockdown [32]. On 27 December 2020, Italy launched the vaccination campaign, which has significantly reduced the risk of COVID-19 diagnosis and COVID-19-related hospitalization and death, particularly starting 14 days from receipt of the second dose; however, the emergence of new SARS CoV-2 variants underscores the importance of receiving a third dose of COVID-19 vaccine to protect the high-risk populations (i.e., older adults) [33].

Despite the rapid development of safe and effective vaccines, there is a need to reduce viral replication, especially if new variants are associated with higher rates of (re)infection or more severe disease. An ever-increasing volume of epidemiological data offers the opportunity to apply data analytics methodologies to extract useful models able to automatically detect both which areas of the country have the highest diagnoses and how the transmission rate of each specific area varies with respect to the time period. This knowledge allows us to dynamically monitor the spread of the disease, offering the opportunity to make better policies to overcome the problem.

Overall, there are three main types of statistical modeling used for predicting infectious disease spread [34]. The first one among them is the *distribution fitting* technique, wherein most of the infectious diseases as a large number of cases are infected is fitted to the observed data and the parameters of the distribution are estimated based on the sample observations. For example, Hamzaha et al. [35] analyzed worldwide COVID-19 data to predict new cases, deaths, and recoveries using distribution methods.

A second type of infectious disease modeling includes *epidemiological models* (e.g., SIR and SEIR), which aim to describe, analyze, and understand the patterns of infectious disease. During the COVID-19 pandemic, several works investigated the effects of non-pharmaceutical measures (such as school closures, travel bans, and national lockdowns) on the spread of COVID-19. Specifically, extensions of well-established Susceptible–Infectious–Recovered (SIR) and Susceptible–Exposed–Infectious–Recovered (SEIR) models have been proposed to model the spread of COVID-19 [36–38]. Such analyses are typically based on complex compartmental models, which focus on individual-level dynamics.

This is a rather different approach as compared to our method, which uses the *time series modeling* technique, as detailed described in Section 2. In particular, we defined a general methodology to estimate and visualize differences in the spread of the pandemic at the national level, with a predefined (but extensible) set of steps (i.e., data collection, pre-processing, analysis, and visualization). In this way, data scientists and analysts can

efficiently design and execute their applications dealing with epidemiological data. In fact, an important advantage of the work is that the users can download the Python code to reproduce the results presented in this paper and monitor the evolution of the pandemic. Moreover, the code could be adapted without much work to monitor the COVID-19 pandemic in other regions, or for future outbreaks of other infectious diseases.

The potential limitations of our study are the following: (i) the data provided by the Department of Civil Protection refer to the number of reported cases, which underestimates the real number of positive cases in the population; (ii) the data do not allow us to ascertain the date of onset of the infection, so the model will suffer from a delay relative to the trend of the infections in the population; (iii) the SARIMA model used, contrary to SARIMAX, does not take into account exogenous variables that could have effects both on the trend and seasonality; (iv) our model is limited to short-term prediction because, with current data, the only observable seasonality has a period of 7 days; (v) this model does not take into account the effect of case spikes due to sudden mutations of the virus and further research is needed to explore such effects on SARIMA model predictions.

Similar SARIMA models can be used for short-term predictions also in other regions and/or countries [13,14,26] but the best model, optimizing Bayesian Information Criterion (BIC) or Akaike Information Criterion (AIC), should be chosen based on available data. A similar model is currently used on the website [epidata.it](https://www.epidata.it/Italia/ARIMA.html) to forecast the next 14 days of new daily COVID-19 cases in Italy <https://www.epidata.it/Italia/ARIMA.html>, accessed on 28 May 2022.

Thus, this study confirms that SARIMA models can be used for short-term predictions (14 days) considering the circaseptan oscillations as seasonality, even at regional level. The best model needs to be chosen via a grid-search to optimize BIC or AIC based on available data.

5. Conclusions

Experimental evaluation, focusing on Calabria, showed a good predictive power for out-of-sample predictions within one week ($R^2 > 0.7$), whereas the predictions up to 14 days should be treated with caution since the predictive power decreases with increasing out-of-sample forecast periods. In future work, other research issues may be investigated. First, we may identify potential exogenous variables to define a better performing SARIMAX model on medium-to-long-term predictions. Second, we will extend the use of the model at the province level, for quickly identifying the potential risk areas of a region, as well as to explore the use of these models to predict the trend of other kinds of events (e.g., deaths and hospitalizations).

Author Contributions: All the authors contributed to the structuring of this paper. F.B. designed the methodology and performed experimental evaluations. M.P. developed the model and performed numerical simulations. S.M. and L.A. reviewed the content of the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pneumonia of Unknown Cause—China. Available online: <https://www.who.int/csr/don/05-january-2020-pneumonia-of-unknown-cause-china/en/> (accessed on 22 March 2022).
2. Listings of WHO's Response to COVID-19. Available online: <https://www.who.int/news-room/detail/29-06-2020-covidtimeline> (accessed on 20 March 2022).

3. Tracking SARS-CoV-2 Variants. Available online: <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/> (accessed on 20 March 2022).
4. Update on Omicron. Available online: <https://www.who.int/news/item/28-11-2021-update-on-omicron> (accessed on 20 March 2022).
5. Statement on Omicron Sublineage, BA.2. Available online: <https://www.who.int/news/item/22-02-2022-statement-on-omicron-sublineage-ba.2> (accessed on 20 March 2022).
6. Four Possible Cases of BA.2.2 Omicron Sub-Variant Detected in Thailand no Cause for Alarm. Available online: <https://www.thaipbsworld.com/four-possible-cases-of-ba.2-2-sub-variant-detected-in-thailand-no-cause-for-alarm/> (accessed on 20 March 2022).
7. Adeola, A.M.; Botai, J.O.; Rautenbach, H.; Adisa, O.M.; Ncongwane, K.P.; Botai, C.M.; Adebayo-Ojo, T.C. Climatic variables and malaria morbidity in mutale local municipality, South Africa: A 19-year data analysis. *Int. J. Environ. Res. Public Health* **2017**, *14*, 1360. [CrossRef] [PubMed]
8. Choi, S.B.; Ahn, I. Forecasting seasonal influenza-like illness in South Korea after 2 and 30 weeks using Google Trends and influenza data from Argentina. *PLoS ONE* **2020**, *15*, e0233855. [CrossRef] [PubMed]
9. He, J.; He, J.; Han, Z.; Teng, Y.; Zhang, W.; Yin, W. Environmental Determinants of Hemorrhagic Fever with Renal Syndrome in High-Risk Counties in China: A Time Series Analysis (2002–2012). *Am. J. Trop. Med. Hyg.* **2018**, *99*, 1262. [CrossRef] [PubMed]
10. Watad, A.; Watad, S.; Mahroum, N.; Sharif, K.; Amital, H.; Bragazzi, N.L.; Adawi, M. Forecasting the West Nile virus in the United States: An extensive novel data streams-based time series analysis and structural equation modeling of related digital searching behavior. *JMIR Public Health Surveill.* **2019**, *5*, e9176. [CrossRef]
11. Duan, Y.; Huang, X.L.; Wang, Y.J.; Zhang, J.Q.; Zhang, Q.; Dang, Y.W.; Wang, J. Impact of meteorological changes on the incidence of scarlet fever in Hefei City, China. *Int. J. Biometeorol.* **2016**, *60*, 1543–1550. [CrossRef]
12. Zhao, Y.; Li, R.; Qiu, J.; Sun, X.; Gao, L.; Wu, M. Prediction of human brucellosis in China Based on temperature and NDVI. *Int. J. Environ. Res. Public Health* **2019**, *16*, 4289. [CrossRef]
13. Chaurasia, V.; Pal, S. COVID-19 pandemic: ARIMA and regression model-based worldwide death cases predictions. *SN Comput. Sci.* **2020**, *1*, 1–12.
14. Tan, C.V.; Singh, S.; Lai, C.H.; Zamri, A.S.S.M.; Dass, S.C.; Aris, T.B.; Ibrahim, H.M.; Gill, B.S. Forecasting COVID-19 Case Trends Using SARIMA Models during the Third Wave of COVID-19 in Malaysia. *Int. J. Environ. Res. Public Health* **2022**, *19*, 1504. [CrossRef]
15. Cori, A.; Ferguson, N.M.; Fraser, C.; Cauchemez, S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.* **2013**, *178*, 1505–1512.
16. Durbin, J.; Koopman, S.J. *Time Series Analysis by State Space Methods*; OUP Oxford: Oxford, UK, 2012; Volume 38.
17. De Gooijer, J.G.; Abraham, B.; Gould, A.; Robinson, L. Methods for determining the order of an autoregressive-moving average process: A survey. *Int. Stat. Rev. Int. Stat.* **1985**, *53*, 301–329. [CrossRef]
18. Elliott, G.; Rothenberg, T.J.; Stock, J.H. *Efficient Tests for an Autoregressive Unit Root*; NBER: Cambridge, MA, USA, 1992.
19. Italian COVID-19 Data Repository. Available online: <https://github.com/pcm-dpc/COVID-19> (accessed on 25 March 2022).
20. From Infection Report to Vaccines: All DATA on the Covid Emergency in Calabria on a Single Platform. Available online: <https://www2.unical.it/portale/portaltemplates/view/view.cfm?109945> (accessed on 25 March 2022).
21. Bonifazi, G.; Lista, L.; Menasce, D.; Mezzetto, M.; Pedrini, D.; Spighi, R.; Zoccoli, A. A simplified estimate of the effective reproduction number R_t using its relation with the doubling time and application to Italian COVID-19 data. *Eur. Phys. J. Plus* **2021**, *136*, 1–14. [CrossRef] [PubMed]
22. Cereda, D.; Tirani, M.; Rovida, F.; Demicheli, V.; Ajelli, M.; Poletti, P.; Trentini, F.; Guzzetta, G.; Marziano, V.; Barone, A.; et al. The early phase of the COVID-19 outbreak in Lombardy, Italy. *arXiv* **2020**, arXiv:2003.09320.
23. Pavlíček, T.; Rehak, P.; Král, P. Oscillatory dynamics in infectivity and death rates of COVID-19. *Msystems* **2020**, *5*, e00700–20. [CrossRef] [PubMed]
24. Huang, J.; Liu, X.; Zhang, L.; Zhao, Y.; Wang, D.; Gao, J.; Lian, X.; Liu, C. The oscillation-outbreaks characteristic of the COVID-19 pandemic. *Natl. Sci. Rev.* **2021**, *8*, nwab100. [CrossRef]
25. Bukhari, Q.; Jameel, Y.; Massaro, J.M.; D’Agostino, R.B.; Khan, S. Periodic oscillations in daily reported infections and deaths for coronavirus disease 2019. *JAMA Netw. Open* **2020**, *3*, e2017521. [CrossRef]
26. Arunkumar, K.; Kalaga, D.V.; Kumar, C.M.S.; Chilkoor, G.; Kawaji, M.; Brenza, T.M. Forecasting the dynamics of cumulative COVID-19 cases (confirmed, recovered and deaths) for top-16 countries using statistical machine learning models: Auto-Regressive Integrated Moving Average (ARIMA) and Seasonal Auto-Regressive Integrated Moving Average (SARIMA). *Appl. Soft Comput.* **2021**, *103*, 107161.
27. Knaub, J.R., Jr. Essential Heteroscedasticity. 2017. Available online: https://www.researchgate.net/publication/32853387_Essential_Heteroscedasticity (accessed on 28 May 2022).
28. Box, G.E.; Cox, D.R. An analysis of transformations. *J. R. Stat. Soc. Ser. B (Methodol.)* **1964**, *26*, 211–243. [CrossRef]
29. Nontapa, C.; Kesamoon, C.; Kaewhawong, N.; Intrapai boon, P. A New Time Series Forecasting Using Decomposition Method with SARIMAX Model. In Proceedings of the International Conference on Neural Information Processing, Bangkok, Thailand, 18–22 November 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 743–751.

30. Abenavoli, L.; Cinaglia, P.; Luzzza, F.; Gentile, I.; Boccutto, L. Epidemiology of coronavirus disease outbreak: The Italian trends. *Rev. Recent Clin. Trials* **2020**, *15*, 87–92. [[CrossRef](#)]
31. Abenavoli, L.; Cinaglia, P.; Procopio, A.C.; Serra, R.; Aquila, I.; Zanza, C.; Longhitano, Y.; Artico, M.; Larussa, T.; Boccutto, L.; et al. SARS-CoV-2 Spread Dynamics in Italy: The Calabria Experience. *Rev. Recent Clin. Trials* **2021**, *16*, 309–315. [[CrossRef](#)]
32. Guzzetta, G.; Riccardo, F.; Marziano, V.; Poletti, P.; Trentini, F.; Bella, A.; Andrianou, X.; Del Manso, M.; Fabiani, M.; Bellino, S.; et al. Impact of a nationwide lockdown on SARS-CoV-2 transmissibility, Italy. *Emerg. Infect. Dis.* **2021**, *27*, 267. [[CrossRef](#)]
33. Branda, F. Impact of the additional/booster dose of COVID-19 vaccine against severe disease during the epidemic phase characterized by the predominance of the Omicron variant in Italy, December 2021–May 2022. *medRxiv* **2022**. [[CrossRef](#)]
34. Yadav, S.K.; Akhter, Y. Statistical Modeling for the Prediction of Infectious Disease Dissemination With Special Reference to COVID-19 Spread. *Front. Public Health* **2021**, 680. [[CrossRef](#)] [[PubMed](#)]
35. Hamzah, F.B.; Lau, C.; Nazri, H.; Ligot, D.V.; Lee, G.; Tan, C.L.; Shaib, M.; Zaidon, U.H.B.; Abdullah, A.B.; Chung, M.H.; et al. CoronaTracker: Worldwide COVID-19 outbreak data analysis and prediction. *Bull. World Health Organ.* **2020**, *1*, 1–32.
36. Flaxman, S.; Mishra, S.; Gandy, A.; Unwin, H.J.T.; Mellan, T.A.; Coupland, H.; Whittaker, C.; Zhu, H.; Berah, T.; Eaton, J.W.; et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* **2020**, *584*, 257–261. [[CrossRef](#)] [[PubMed](#)]
37. Giordano, G.; Blanchini, F.; Bruno, R.; Colaneri, P.; Di Filippo, A.; Di Matteo, A.; Colaneri, M. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat. Med.* **2020**, *26*, 855–860. [[CrossRef](#)]
38. Lin, Q.; Zhao, S.; Gao, D.; Lou, Y.; Yang, S.; Musa, S.S.; Wang, M.H.; Cai, Y.; Wang, W.; Yang, L.; et al. A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. *Int. J. Infect. Dis.* **2020**, *93*, 211–216. [[CrossRef](#)]