



# Forecasting the number of confirmed new cases of COVID-19 in Italy for the period from 19 May to 2 June 2020



Marco Triacca <sup>a</sup>, Umberto Triacca <sup>b,\*</sup>

<sup>a</sup> Department of Mathematics, Guido Castelnuovo, University of Rome “La Sapienza”, Italy

<sup>b</sup> Department of Information Engineering, Computer Science and Mathematics, University of L'Aquila, Italy

## ARTICLE INFO

### Article history:

Received 5 October 2020

Received in revised form 2 January 2021

Accepted 10 January 2021

Available online 26 January 2021

Handling editor: Dr. J Wu

### Keywords:

COVID-19

Real-time forecasts

Time series

## ABSTRACT

In this paper we forecast the spread of the coronavirus disease 2019 outbreak in Italy in the time window from May 19 to June 2, 2020. In particular, we consider the forecast of the number of new daily confirmed cases. A forecast procedure combining a log-polynomial model together with a first-order integer-valued autoregressive model is proposed. An out-of-sample comparison with forecasts from an autoregressive integrated moving average (ARIMA) model is considered. This comparison indicates that our procedure outperforms the ARIMA model. The Root Mean Square Error (RMSE) of the ARIMA is always greater than that of the our procedure and generally more than twice as high as the our procedure RMSE. We have also conducted Diebold and Mariano (1995) tests of equal mean square error (MSE). The tests results confirm that forecasts from our procedure are significantly more accurate at all horizons. We think that the advantage of our approach comes from the fact that it explicitly takes into account the number of swabs.

© 2021 The Authors. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Coronavirus disease 2019 (COVID-19) is defined as illness caused by a novel coronavirus now called “severe acute respiratory syndrome coronavirus 2” (SARSCoV-2; formerly called 2019-nCoV), which was first identified amid an outbreak of respiratory illness cases in Wuhan City, Hubei Province, China. The World Health Organization declared COVID-19 a pandemic on March 11, 2020.

Public health officials are increasingly using mathematical, statistical and computational tools to guide intervention plans against infectious disease outbreaks. In particular, when an outbreak occurs, statistical models can be useful to forecast the trajectory of the outbreak. In this paper, we propose a simple econometric procedure to predict the spread of COVID-19. The forecasts are obtained in the time window from May 19 to June 2, 2020, in Italy.

A number of authors has tried to forecast the evolution of the COVID-19 outbreak by using time series methods; these include classical ARIMA models, (Alzahrani et al. (2020), Benvenuto et al. (2020), Chintalapudi et al. (2020), Perone (2020); Yang et al. (2020)), a hybrid ARIMA-Wavelet-based model (Chakraborty and Ghosh (2020)), a log-polynomial model (Peracchi

\* Corresponding author.

E-mail address: [umberto.triacca@univaq.it](mailto:umberto.triacca@univaq.it) (U. Triacca).

Peer review under responsibility of KeAi Communications Co., Ltd.

(2020)), machine learning models (Dal Molin Ribeiro et al. (2020)) and artificial neural network models (Wieczorek et al. (2020)).

In order to forecast the number of daily new diagnosed cases, we use a log-polynomial model together with a first-order INteger-valued AutoRegressive (*INAR*(1)) model. The log-polynomial model is used to forecast the ratio *number of daily new diagnosed cases/number of swabs* and the *INAR*(1) model is used to forecast the *number of swabs*. The forecast of the number of daily new diagnosed cases is then obtained by multiplying these two forecasts. An out-of-sample comparison with forecasts from an ARIMA model is also considered. Our procedure appears to be providing a very good forecast performance. Probably, the advantage of our approach comes from the fact that it explicitly takes into account the number of swabs. In fact, we think that it is necessary to consider the number of swabs to capture the fluctuations of the number of daily new confirmed cases around the trend. To the best of our knowledge, no model with this characteristic is present in literature.

The structure of the paper is as follows. Section 2 describes the used data set. Section 3 illustrates the used statistical models. Section 4 presents real-time forecasts of confirmed cases. Section 5 concludes the paper.

## 2. Characteristics of the data

In this paper we use daily time-series data from the Italian Civil Protection Department (Dipartimento della Protezione Civile or DPC). Data are available at: <https://github.com/pcm-dpc/COVID-19/tree/master/dati-andamentonazionale>. The DPC website contains different daily time-series at various levels of aggregation, starting February 24, 2020. For a complete description of this data set, see Peracchi (2020). Here, we consider two time series, at national level, from March 18 to June 2, 2020:

- the number of new daily COVID-19 confirmed cases (“Nuovi positivi”);
- the number of daily nasal swabs (“Tamponi”).

Fig. 1 illustrates the behavior of the new daily confirmed cases. This series seems to fluctuate around a linear time trend.

The number of daily swab tests in Fig. 2 shows significant weekly fluctuations too. It is interesting to note that there is a strong positive linear relationship between “Nuovi positivi” and “Tamponi” (see the scatter plot in Fig. 3). The correlation coefficient between these variables (0.88) is significantly different from zero.

This correlation suggests that a model that does not take into account the number of performed swabs might not provide a good predictive performance. In order to address this problem, we adopt the following procedure. First, we use a log-polynomial model to forecast the ratio Nuovi Positivi/Tamponi, then we use an *INAR*(1) model to forecast the number of Tamponi, and finally we combine the two forecasts.

## 3. The statistical models

In this section we introduce, shortly, the log-polynomial and *INAR*(1) models and we present our forecast procedure.

### 3.1. The log-polynomial model

Let  $Y_t$  be a non-negative integer-valued random variable. We say that  $Y_t$  follows a log-polynomial model of order  $n$  if

$$\ln(Y_t) = \beta_0 + \beta_1 t + \dots + \beta_n t^n + u_t, \quad t = 1, 2, \dots \quad (1)$$

where  $\beta_0, \beta_1, \dots, \beta_n$  are  $n+1$  unknown parameters and  $u_t$  is a random term, with mean zero and variance  $\sigma^2$ .

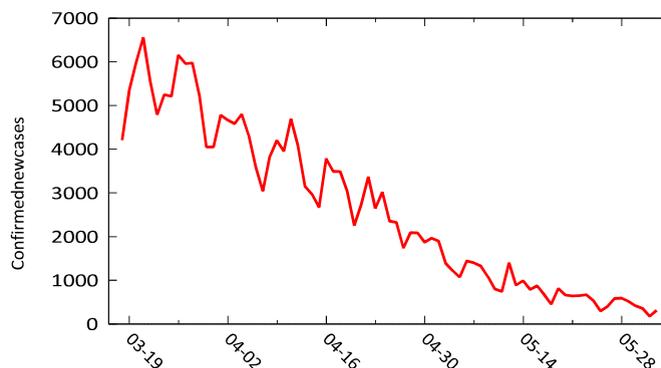


Fig. 1. Number of new daily COVID-19 confirmed cases from March 18 to June 2, 2020

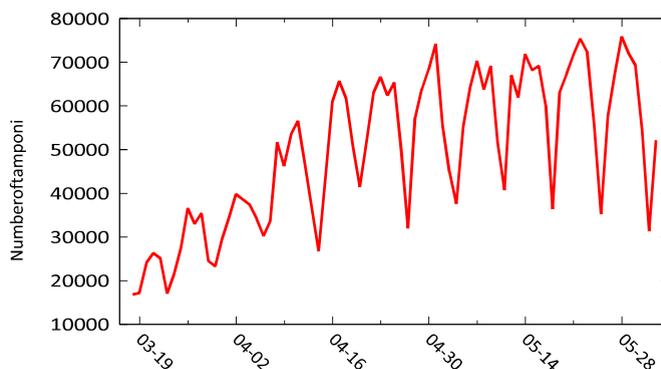


Fig. 2. Number of Tamponi from March 18 to June 2, 2020

To fix ideas, we consider the case  $n = 2$ . If we let the historical data be denoted by  $Y_1, \dots, Y_T$ , then we can write the forecasts of  $Y_{T+h}$  for  $h = 1, 2, \dots$  as

$$\tilde{Y}_{T+h|T} = \exp(\beta_0 + \beta_1(T+h) + \beta_2(T+h)^2) \quad h = 1, 2, \dots$$

It is important to note that the predictor  $\tilde{Y}_{T+h|T}$  will systematically underestimate the conditional mean function

$$\mathbb{E}[Y_{T+h}|X_T, X_{T-1}, \dots] = \exp(\beta_0 + \beta_1(T+h) + \beta_2(T+h)^2) \mathbb{E}[\exp(u_{T+h})]$$

In fact, the Jensen inequality, together with the condition  $\mathbb{E}[u_t] = 0$ , implies that

$$\mathbb{E}[\exp(u_{T+h})] > 1.$$

Duan (1983) suggested that a consistent estimator of  $\mathbb{E}[\exp(u_{T+h})]$  is given by

$$\hat{\alpha}_0 = \frac{1}{T} \sum_{t=1}^T \exp(\hat{u}_t)$$

where  $\hat{u}_t$  is a least squares residual in the original log form regression (1). Thus an appropriate feasible predictor for  $Y_{t+h}$  would be the so-called Duan's smearing estimator

$$\hat{Y}_{T+h|T} = \exp(\hat{\beta}_0 + \hat{\beta}_1(T+h) + \hat{\beta}_2(T+h)^2) \hat{\alpha}_0 \tag{2}$$

where  $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$  are the Ordinary Least Square Estimates of the parameters  $\beta_0, \beta_1, \beta_2$ , respectively. In the sequel we will use this predictor.

### 3.2. The INAR(1) model

The INAR(1) process was introduced by Mckenzie, 1988, Al-Osh and Alzaid (1987) and Alzaid and Al-Osh (1988) in order to modelling non-negative integervalued time series.

Before introducing the INAR(1) model, it is necessary to give the definition of the thinning operator, as reported by Steutel and van Ham, 1979.

Let  $X$  be a non-negative integer-valued random variable and  $\alpha \in [0,1]$ , the binomial thinning operator, denoted by  $\circ$ , is defined as

$$\alpha \circ X = \sum_{j=1}^X V_j$$

where  $\{V_j; j = 1, \dots, X\}$  is a sequence of independent and identically distributed (i.i.d.) random variables that follow a Bernoulli distribution with parameter  $\alpha$ , and that is independent of  $X$ .

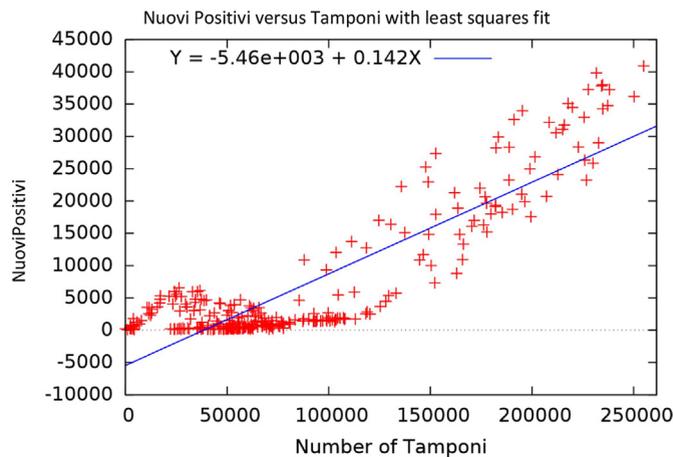


Fig. 3. Scatter plot Nuovi positivi versus Tamponi, February 24 to December 25, 2020

A discrete-time non-negative integer-valued stochastic process  $\{X_t; t \in \mathbb{Z}\}$  is said to be an  $INAR(1)$  process if it satisfies the equation

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t \quad t \in \mathbb{Z}$$

where  $\{\varepsilon_t; t = 1, 2, \dots\}$  is an innovation sequence of i.i.d non-negative integer-valued random variables not depending on past values of  $\{X_t; t \in \mathbb{Z}\}$ .

Following Bourguignon et al. (2016) we also introduce the definition of first order seasonal  $INAR$  process.

A discrete-time non-negative integer-valued stochastic process  $\{X_t; t \in \mathbb{Z}\}$  is said to be a first-order seasonal  $INAR$  process with seasonal period  $s$  ( $INAR(1)_s$ ) if it satisfies the following equation:

$$X_t = \alpha \circ X_{t-s} + \varepsilon_t \quad t \in \mathbb{Z}$$

where  $\{\varepsilon_t; t \in \mathbb{Z}\}$  is an innovation sequence of i.i.d non-negative integer-valued random variables not depending on past values of  $\{X_t; t \in \mathbb{Z}\}$  and  $s \in \mathbb{N}$  denotes the seasonal period. If  $\{\varepsilon_t; t \in \mathbb{Z}\}$  is a sequence of independent and identically distributed Poisson random variables with mean  $\lambda$ , then the process  $\{X_t; t \in \mathbb{Z}\}$  is called a Poisson  $INAR(1)_s$  process.

In the sequel we will use a Poisson  $INAR(1)_s$  process. The unknown parameters of this model are  $\alpha$  and  $\lambda$ . We can estimate  $\alpha$  and  $\lambda$  using the conditional least squares (CLS) method.

Let  $X_0, X_1, \dots, X_T$  be a sample of a Poisson  $INAR(1)_s$  process. For all  $k \in \mathbb{Z}_+$ ,  $F_k$  denotes the  $\sigma$ -algebra generated by the random variables  $X_0, X_1, \dots, X_k$  and  $\mathbb{E}_\theta$  denotes the expectation with respect to the true parameter  $\theta = (\alpha, \lambda)'$ .

The CLS estimator  $\hat{\theta}_{CLS} = (\hat{\alpha}_{CLS}, \hat{\lambda}_{CLS})$  of  $\theta = (\alpha, \lambda)'$  is given by

$$\hat{\theta}_{CLS} = \operatorname{argmin}_\theta (Q_T(\theta))$$

where  $Q_T(\theta) = \sum_{t=s+1}^T [X_t - \mathbb{E}(X_t|F_{t-1})]^2$ . It is possible to show that the CLS estimates of  $\alpha$  and  $\lambda$  are given, respectively, as

$$\hat{\alpha}_{CLS} = \frac{(T-s) \sum_{t=s+1}^T X_t X_{t-s} - \sum_{t=s+1}^T X_t \sum_{t=s+1}^T X_{t-s}}{(T-s) \sum_{t=s+1}^T X_{t-s}^2 - \left( \sum_{t=s+1}^T X_{t-s} \right)^2}$$

$$\hat{\lambda}_{CLS} = \frac{1}{T-s} \left( \sum_{t=s+1}^T X_t - \hat{\alpha}_{CLS} \sum_{t=s+1}^T X_{t-s} \right)$$

Now, we consider the problem of forecasting a future value  $X_{T+h}$ ,  $h \in \mathbb{N}$ , based on the series up to time  $T \in \mathbb{N}$ . Following Bourguignon et al. (2016) we will use the predictor

$$\hat{X}_{T+h|T} = \hat{\alpha}_{CLS}^q \left( X_{T-r} - \frac{\hat{\lambda}_{CLS}}{1 - \hat{\alpha}_{CLS}} \right) + \frac{\hat{\lambda}_{CLS}}{1 - \hat{\alpha}_{CLS}} \quad h \in \mathbb{N} \tag{3}$$

where  $q = [h/s]$  and  $r = qs - h$ , with  $[x]$  denoting the upper integer part of  $x \in \mathbb{R}$ , that is,  $[x] = \min\{T \in \mathbb{N} | x \leq T\}$ .

### 3.3. A procedure to forecast the confirmed cases of COVID19

In this subsection we illustrate our procedure to forecast the number of daily new infections. We denote with  $Z_t$  the variable new infections and with  $X_t$  the number of swabs. Then, we consider the ratio  $Y_t = Z_t/X_t$ .

Our predictor (denoted by  $\widehat{Z}_{T+h|T}$ ) is obtained in three steps:

**Step 1.** To estimate a log-polynomial model of order 1 for the variable  $Y_t$  and to use the predictor (2) to forecast  $Y_{T+h}$ ,  $h = 1, 2, \dots$

**Step 2.** To estimate an  $INAR_7(1)$  for the variable  $X_t$  and to use the predictor (3) to forecast  $X_{T+h}$ ,  $h = 1, 2, \dots$

**Step 3.** To obtain  $\widehat{Z}_{T+h|T}$  by

$$\widehat{Z}_{T+h|T} = \widehat{Y}_{T+h|T} \widehat{X}_{T+h|T} \quad h = 1, 2, \dots$$

To the best of our knowledge, this is the first paper presenting such a combination of two different approaches (log-polynomial and  $INAR(1)$  models) to forecast the confirmed cases of COVID-19.

## 4. Real-time forecasting performance of the proposed procedure

In this section we evaluate the real-time forecasting performance of our procedure and compare it with the forecasts from an ARIMA model for the number of confirmed new cases. We denote the forecasts from our procedure with P1 and the ARIMA forecasts with P2.

We remember that a variable  $y$  follows an  $ARIMA(p, d, q)$  if

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)(1 - L)^d y_t = (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) u_t$$

where  $L$  is the lag operator,  $d$  is the degree of first differencing involved,  $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$  and  $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q$  are the autoregressive and moving average polynomials of orders  $p$  and  $q$  respectively,  $u_t$  is a normally distributed white noise process with mean 0 and variance  $\sigma_u^2$ .

ARIMA models often outperform more sophisticated structural models in terms of short-run forecasting ability. Therefore, the ARIMA forecasting technique provides a useful benchmark by which other forecasting techniques may be appraised. In particular, we have identified (using the BIC criterion) and estimated an  $ARIMA(2,1,2)$  for the number of confirmed new cases.<sup>1</sup>

In our real-time forecast evaluation exercise we use a recursive forecasting scheme, expanding the model estimation sample before each new forecast. In the first run the models are estimated on the period March 18 - May 18, 2020 (training set) and the evaluation period is May 19 - June 2, 2020 (test set), in the second run the models are estimated on the period March 18 - May 19, 2020 and the test set is May 20 - June 2, 2020, and so forth. In this way we obtain, for each procedure P1 and P2, 15 one-day-ahead predictions, 14 two-day-ahead predictions, 13 three-day-ahead predictions, and so on.

### 4.1. Motivation of the time span March 18 - June 2 2020

It is important to underline that our procedure, like many other time series methods, assumes that the underlying data generating process of the time series is constant. For actual time series data this assumption is often invalid as shifting environmental conditions may cause the underlying data generating process to change. In the period March 18 - June 2, 2020 (within which both the training sets and test sets are included) the assumption of constancy of the data generating process can be considered a plausible assumption, because Italy went into lockdown from March 9 to June 3, when the Italian Government granted free movement to all citizens across Regions.

### 4.2. Performance measures

We follow the usual practice in the literature, and evaluate the point forecasts from the different models using both the square root of the mean squared forecast errors (RMSE) and the mean of the absolute forecast errors (MAE). In algebraic terms they can be defined as, respectively,

<sup>1</sup> The parameters of the ARIMA models are significantly different from zero and the LjungBox Q-test does not reject the null hypothesis that the series of residuals exhibits no autocorrelation.

$$RMSE(h) = \sqrt{\frac{1}{N - (h - 1)} \sum_{i=h}^N \left( W_{T+i} - \widehat{W}_{T+i|T-h+i} \right)^2}, \quad h = 1, 2, \dots, P$$

and

$$MAE(h) = \frac{1}{N - (h - 1)} \sum_{i=h}^N \left| W_{T+i} - \widehat{W}_{T+i|T-h+i} \right|, \quad h = 1, 2, \dots, P$$

where  $\widehat{W}_{T+i|T-h+i}$  is the predicted value of  $W_{T+i}$  based on the estimates of the model parameters from all the data available up to the current date  $T - h + i$ ,  $N$  is a positive integer, and  $0 < P < N$ .

These measures capture the average deviation of the forecast from the realized value over the training set.

Table 1 shows the RMSE and the MAE for the two forecasting procedures (P1 and P2). The procedure with the smallest forecasting error for each horizon is displayed in bold. We note that the proposed procedure P1 always outperforms the alternative ARIMA-based procedure P2. It is evident that the RMSE of the procedure P2 is always greater than that of the Procedure P1 and generally more than twice as high as the P1 RMSE.

We have also conducted Diebold and Mariano (1995) tests of equal mean square error (MSE). The results of the tests (available upon request) confirm that forecasts from procedure P1 are significantly more accurate at all horizons.

In particular, the good performance of the procedure P1, when we consider the forecast one step ahead, is shown in Fig. 4.

We close this section considering Fig. 5. It presents a comparison between the actual number of confirmed new cases and the predicted values obtained from P1 procedure and P2 procedure.

Here, all the forecasts are performed at time May 18 for the period from May 19 to June 2, 2020. It is evident that the forecasts from procedure P1 match very well the actual number of confirmed new cases, while the forecasts from P2 procedure are characterized by a larger deviation.

### 5. Conclusions and future work

This paper has aimed to provide a forecast of the COVID-19 diffusion in Italy by using a data driven approach. In particular, we have proposed a new procedure based on the combined use of a log-polynomial model and an *INAR*(1) model. We have used this procedure to forecast the number of new daily confirmed cases in the time window from May 19 to June 2, 2020. An out-of-sample comparison with forecasts from an autoregressive integrated moving average (ARIMA) model has been considered. This comparison indicates that our procedure outperforms the ARIMA model. The accuracy of the short-term forecasts from our procedure is very high on the considered period. We think that the advantage of our approach comes from the fact that it explicitly takes into account the number of swabs.

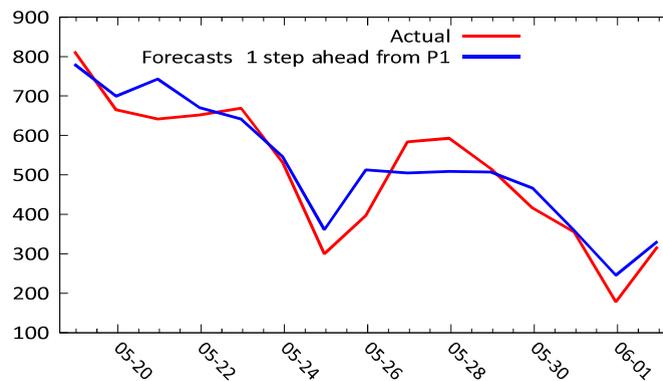
The accuracy of the predictions could be augmented including the reproduction number values ( $R_t$ ) in our procedure. Unfortunately, we think that this is not possible as our procedure combines two models that do not allow the introduction of other variables different from the numbers of swabs ( $X_t$ ) and the ratio new daily confirmed cases/numbers of swabs ( $Y_t$ ). This happens because  $X_t$  is modeled by an autoregression and  $Y_t$  is considered a function of the time only. Maybe a possibility could be to substitute the log-polynomial model for  $Y_t$  with a bivariate Vector AutoRegressive (VAR) model modeling  $Y_t$  and  $R_t$ . We will consider this approach in future work. Another point that we intend to investigate in future research concerns the comparison of the predictive performance of our procedure with a forecasting model that includes the reproduction number values (examples of such models can be found in Chintalapudi et al. (2020b) and Zhao et al. (2020)).

We hope that the results of our analysis contribute to the elucidation of critical aspects of this outbreak providing a useful perspective in Italy and internationally on how the pandemic spreads. In particular, forecasts can be helpful to anticipate resources such as protective equipment, medical ventilators, and hospital beds.

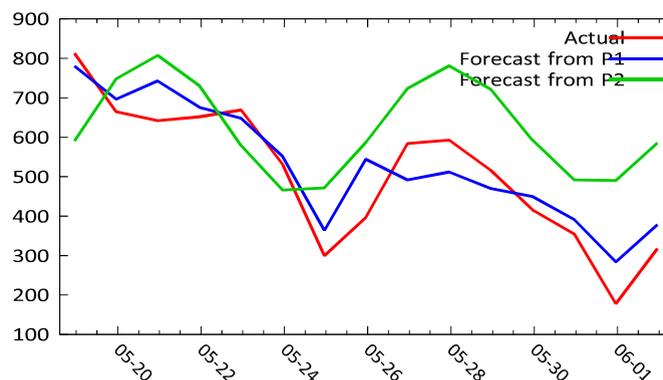
**Table 1**  
RMSE and MAE of the forecasts.

RMSE							
	h = 1	h = 2	h = 3	h = 4	h = 5	h = 6	h = 7
P1	<b>58.66</b>	<b>60.15</b>	<b>60.48</b>	<b>55.54</b>	<b>58.89</b>	<b>62.13</b>	<b>64.64</b>
P2	121.95	127.54	115.79	103.60	130.95	179.50	197.33
MAE							
	h = 1	h = 2	h = 3	h = 4	h = 5	h = 6	h = 7
P1	<b>47.56</b>	<b>49.01</b>	<b>49.47</b>	<b>44.38</b>	<b>47.71</b>	<b>50.54</b>	<b>53.05</b>
P2	92.27	95.82	95.57	85.34	114.56	141.18	166.52

Note. The table shows the RMSE and MAE of the forecasts obtained from P1 and P2 procedures. The values in bold represent the best model according to each measure of error and for each forecast horizon.



**Fig. 4.** Forecasts of confirmed new cases 1 step ahead from P1 procedure, from May 19 to June 2, 2020.



**Fig. 5.** Forecasts of confirmed new cases from P1 and P2 (ARIMA) procedures, from May 19 to June 2, 2020. All the forecasts are performed at time May 18.

Finally, it is important to note that a limitation of our analysis is that it has been conducted only using the reported number of confirmed new cases that have been officially notified. The current numbers of confirmed cases are known to be vastly underestimated due to the limited testing available. There are almost certainly many more cases of COVID-19 that have not been diagnosed than those that have. Nevertheless, while these data are biased, they are the only source of information available that can guide our efforts to understand the outbreak dynamics.

### Declaration of competing interest

The authors declare that they have no conflict of interest.

### Acknowledgements

The authors would like to thank Fulvia Focker and Sergio Vitale for their comments. We also thank the anonymous reviewer for its suggestions.

### References

- Al-Osh, M. A., & Alzaid, A. A. (1987). First-order integer-valued autoregressive ( $INAR(1)$ ) process. *J Time Ser Anal*, 8, 261–275.
- Alzahrani, S. I., Aljamaan, I. A., & Al-Falih, E. A. (2020). Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions. *Journal of Infection and Public Health*, 13, 914–991.
- Alzaid, A. A., & Al-Osh, M. A. (1988). First-order integer-valued autoregressive process: Distributional and regression properties. *Statistica Neerlandica*, 42, 53–61.
- Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., & Ciccozzi, M. (2020). Application of the ARIMA model on the COVID-2019 epidemic data set. <https://doi.org/10.1016/j.dib.2020.105340>. Data in Brief.
- Bourguignon, M., Vasconcellos, K. L. P., Reisen, V. A., & Ispany, M. (2016). A Poisson  $INAR(1)$  process with a seasonal structure. *Journal of Statistical Computation and Simulation*, 86(2), 373–387.
- Chakraborty, T., & Ghosh, I. (2020). Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis. *Chaos, Solitons, and Fractals*, 135, 109850.
- Chintalapudi, N., Battineni, G., & Amenta, F. (2020a). COVID-19 disease outbreak forecasting of registered and recovered cases after sixty day lockdown in Italy: A data driven model approach. *Journal of Microbiology, Immunology, and Infection*, 53(3), 396–403.

- Chintalapudi, N., Battineni, G., Sagaro, G. G., & Amenta, F. (2020b). COVID-19 outbreak reproduction number estimations and forecasting in Marche, Italy. *International Journal of Infectious Diseases*, 96, 327–333.
- Dal Molin Ribeiro, M. H., da Silva, R. G., Mariani, V. C., & dos Santos Coelho, L. (2020). Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. *Chaos, Solitons & Fractals*, 135, 109853. <https://doi.org/10.1016/j.chaos.2020.109853>
- Diebold, F. X., & Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263.
- Duan, N. (1983). Smearing estimate: A nonparametric retransformation method. *Journal of the American Statistical Association*, 78, 605–610.
- Mckenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability*, 20, 822–835.
- Peracchi, F. (2020). The Covid-19 pandemic in Italy. Available at <http://www.eief.it/eief/index.php/forecasts>.
- Perone, G. (2020). *An ARIMA model to forecast the spread and the final size of COVID-2019 epidemic in Italy*. medRxiv.
- Steutel, F. W., & van Ham, K. (1979). Discrete analogues of self-decomposability and stability. *Annals of Probability*, 7, 893–899.
- Wieczorek, M., Sil, ka J., & Woźniak, M. (2020). Neural network powered COVID-19 spread forecasting model. *Chaos, Solitons & Fractals*, 140, 110203. <https://doi.org/10.1016/j.chaos.2020.110203>
- Yang, Q., Wang, J., Ma, H., & Wang, X. (2020). Research on COVID-19 based on ARIMA model-aking Hubei, China as an example to see the epidemic in Italy. *Journal of Infection and Public Health*, 4–7. <https://doi.org/10.1016/j.jiph.2020.06.019>
- Zhao, H., Merchant, N., McNulty, A., Radcliff, T., Cote, M., Fischer, Sang, H., & Ory, M. G. (2020). COVID-19: Short term prediction model using daily incidence data. *The Lancet Public Health*. <https://doi.org/10.2139/ssrn.3727298>. D-20-03662, Available at SSRN <https://ssrn.com/abstract=3727298>.