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.
The growth of COVID-19 in Spain. A view based on time-series forecasting methods

Andrés Carrión-García, José Jabaloyes, Angela Grisales
DEPARTMENT OF APPLIED STATISTICS O.R. AND QUALITY, UNIVERSITAT POLITECNICA DE VALENCIA, VALENCIA, SPAIN

1. Introduction

Since the first cases of what is now known as COVID-19 were detected in China at the end of 2019 [1,2], the Coronavirus disease has spread all over the world, evolving at different speeds and with different patterns, which respond, at least in part, to the different solutions offered by the governments in each country.

In Spain, the first cases were detected in January 2020, and were linked to people arriving from abroad. What initially seemed to be an isolated incident has, in just a few weeks, become the largest major national crisis in recent decades.

Despite having a robust national healthcare system, like other countries, Spain has been fighting the pandemic under highly stressful conditions that have tested its response capacity to the limit, especially in the fastest growing phase of the pandemic’s evolution.

A great international research effort has been developed in the few months after the detection of the new SARS-CoV-2 coronavirus. First papers were centered in the characterization of the virus itself or the pathologies caused. There are many papers linked with these two objectives, and only for illustration purposes, we can mention some of them. Probably one of the first papers is the article by Zhu et al. [1] in which they describe the disease as a new form of pneumonia (with unknown etiology) caused by a novel coronavirus, a problem that seemed to be limited to China, and specifically to the area of Wuhan. As the relevance of the disease was recognized and its international expansion begins, more and more research papers appear in scientific journals worldwide. As soon as in February 2020, the paper by Wu et al. [2] reviews the main researches related with SARS-CoV-2 in the fields of epidemiology, clinical manifestations, diagnosis, treatments, and prevention. Also, some papers, as [3], related this disease with previous coronavirus epidemic episodes as SARS-CoV and MERS-CoV.
As the disease started to be better characterized, research moves to the possibility of vaccine development, as in Refs. [4,5] or [6], and to the search for better medication and treatment, as in Refs. [7–11] or [12].

In the Spanish case, and a part of the researches in the previously commented fields (as [11]), some papers have analyzed specificities of the evolution and situation of the pandemic in Spain. This is the case of papers [13–15], and special interest has the paper by Briz and Serrano [13], who analyze the evolution of the pandemic in Spain from the double spatio-temporal perspective with the goal of exploring the possible existence of a relationship between temperature and the evolution of the accumulated number of COVID-19 cases, a question that has been widely discussed. They use the province as geographic analysis unit and founded no evidence of the mentioned relationship, at least for the period analyzed (the first month of the pandemic). Also, special interest has the paper by Ahmar and Boj [14], who use a model related with those used in the present paper. Their objective is to study the effect of the number of confirmed cases over the Spanish Stock Exchange Index, as a way to forecast, always in a short term, the economic impact of the pandemic.

Some researches have applied ARIMA models to forecast different aspects of the pandemic behavior. We will mention just two examples. The paper by Benevenuto et al. [16] applies ARIMA modeling to dataset covering a period between January and February, and in some moment, they use also the logarithmic transformation, as we will. The paper by Tandon et al. [17] applied ARIMA model to data corresponding to the evolution of the pandemic in India, generating short term forecasts and comparing ARIMA with other forecasting methods.

In this chapter, we have centered our analysis specially in the death toll, as we consider this series more robust than others (as will be commented latter). One of our goals, identifying the peak of the pandemic curve, is a relevant question, which can be answered using different approaches. Some experts have used epidemiological models, such as Susceptible, Infected, Recovered (SIR) or Susceptible, Exposed, Infectious, and Removed or Recovered (SEIR) [18] models yet there are other alternatives, such as the use of time-series analysis models, as shown in this paper. These models have been also used to analyze the relationship between some of the more relevant series describing the evolution of the pandemic: the death toll and the intensive care unit (ICU) demand, and the infected cases and recovered persons.

2. Materials and method

2.1 Time-series models

The problem of forecasting the evolution of natural or social phenomena is an old question that has received numerous responses over time from different viewpoints [19]. The aim of this paper is not to put forward a comprehensive overview of time-series analysis models, though it is relevant to distinguish between causal models and pure
time-series analyses. The first category include [20] methods such as the above-mentioned SIR model, based on considering scientific (in this case, epidemiological) knowledge to build a model that is adjusted to real data by optimizing its parameters. The adequateness of SIR model has been discussed by Postnikov [21].

The second category includes a wide variety of models that take advantage of what we can call inertia in the evolution of time series [22–24]. This inertia is the result of the effect on the measured variable of underlying factors that affect the evolution of the corresponding natural phenomenon. These models do not try to understand this effect, but instead aim to recognize regular patterns in the phenomenon’s evolution and use them to forecast it.

### 2.1.1 ARIMA models

The models used here are a well-known stochastic family of models known as ARIMA models. The acronym ARIMA corresponds to Auto Regressive (AR) Integrated (I) Moving Average (MA). The technique applied was initially developed by Box and Jenkins and is often referred to as the Box-Jenkins Analysis (Box et al. [22], Chapters 3 and 4).

ARIMA models are extremely flexible and can adequately model a wide range of time-series behaviors with just a few parameters [24].

To present the model, it is necessary to define the backshift operator B. This is a symbolic operator that reduces the time index of an element in the series:

\[ B z_t = z_{t-1} \]

The ARIMA model is the combination of three components [22,23]:

- **The AR part:**

\[
(z_t - \varphi_1 z_{t-1} - \varphi_2 z_{t-2} - \ldots - \varphi_p z_{t-p} + a_t)
\]

Using the backshift operator in this expression and moving all terms in z to left side, it can be written as:

\[
(1 - \varphi_1 B - \varphi_2 B^2 - \ldots - \varphi_p B^p) z_t = a_t
\]

\[
\varphi_p(B) z_t = a_t
\]

where \( \varphi_p(B) \) is a polynomial in B of order p, called autoregressive polynomial. It represents the influence of the past p values of the series on its present value (Eq. 35.1) (thus it’s named auto-regressive). The random noise element \( a_t \) is considered to be a white noise process.

- **The MA part of the model:**

\[
z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}
\]

Again, using the backshift operator in this expression, it can be written as:

\[
z_t = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q) a_t
\]

\[
z_t = \theta_q(B) a_t
\]
where now, $\theta_q(B)$ is a polynomial in $B$ of order $q$ (MA polynomial), representing the impact in the present time-series value of the last $q$ random inputs (Eq. 35.2) represented by $a_t$.

- And finally, the Integrated part (I), corresponding to the differences required to stabilize the series (converting the original series into a stationary one)

$$\nabla^d z_t = a_t$$

where $\nabla = (1 - B)$ is called the difference operator by its effect, as it generated the difference between one element of the time series and its previous one.

$$\nabla z_t = z_t - z_{t-1}$$

When all these three elements are put together to obtain an ARIMA model, the abbreviated expression is:

$$\phi_p(B)\nabla^d z_t = \theta_q(B)a_t$$

Or more detailed:

$$(1 - \varphi_1 B - \ldots - \varphi_p B^p)(1 - B)^d z_t = (1 - \theta_1 B - \ldots - \theta_q B^q)a_t$$

(35.3)

And this model is referenced as an ARIMA(p,d,q) model. Orders $p$, $d$, and $q$ may have any integer value, but are only greater than 4 or 5 in a few cases, and they are usually equal to or lower than 2.

There is an extension of ARIMA models that includes the option of cyclical evolution patterns. In this case, completely parallel models can be defined with the specificity of using a new seasonal backshift operator $B^s$, instead of the $B$ backshift operator, where $s$ is the length of the seasonal period (i.e., 7 for daily data with weekly seasonality, 12 for monthly data with yearly seasonality):

$$B^s z_t = z_{t-s}$$

In addition, a new seasonal difference operator is set out as:

$$\nabla_s = 1 - B^s$$

Similar autoregressive and moving average polynomials, now in $B^s$, seasonal backshift operator, can be defined as $\Phi_p(B^s)$ and $\Theta_q(B^s)$, respectively.

The final model, including both seasonal and nonseasonal patterns is represented as the ARIMA(p,d,q) (P,D,Q)s model:

$$\phi_p(B)\Phi_p(B^s)\nabla^d \nabla_s^D z_t = \theta_q(B)\Theta_q(B^s)a_t$$

In our study, only nonseasonal structures were detected, so only nonseasonal ARIMA models have been used. The computation of the model’s parameters requires the use of statistical software. In our case, we used Statgraphics and Matlab.
The forecasting expressions of ARIMA models may have three alternative forms [23]. The first one is simply the development of Eq. (35.3). For instance, for an ARIMA(1,1,1) model, the expression will be:

\[ z_t = (1 + \varphi_1)z_{t-1} - \varphi_1 z_{t-2} + a_t - \theta_1 a_{t-1} \]

In this expression, \( z_i (i < t) \) are the past values of the time series, and \( a_i (i < t) \) are the past model residuals. For projecting the forecasts, values of \( z_i (i > t) \) are substituted by the forecasted values, and \( a_i (i > t) \) are replaced by zero.

The second and third forecasts expressions are the result of moving all polynomials in \( B \) to the same term of the equation. If polynomials are moved to the left side, we refer to this expression as the \( \pi \) weights form:

\[ \frac{\Phi_p(B)\nabla^d}{\theta_q(B)}z_t = a_t \]

And as the result of dividing polynomials is another polynomial, in general of infinite order

\[ (1 - \pi_1 B - \pi_2 B^2 - \pi_3 B^3 - \ldots)z_t = a_t \quad (35.4) \]

And from here:

\[ z_t = \pi_1 z_{t-1} - \pi_2 z_{t-2} - \pi_3 z_{t-3} - \ldots - \pi_i z_{t-i} - \ldots + a_t \]

The decision of truncating this expression can be based on the decreasing values of successive \( \pi \) coefficients.

Similarly, if polynomials are moved to the right side, we refer to this expression as the \( \Psi \) weights form:

\[ z_t = \frac{\theta_q(B)}{\Phi_p(B)\nabla^d}a_t \]

And we can obtain

\[ z_t = a_t - \Psi_1 a_{t-1} - \Psi_2 a_{t-2} - \Psi_3 a_{t-3} - \ldots - \Psi_i a_{t-i} - \ldots \]

Also here, the decision of truncating this expression can be based on the decreasing values of successive \( \Psi \) coefficients.

### 2.1.2 Cross-correlations

The cross-correlation function (CCF) is a way of analyzing the similarity between two time series as a function of the displacement in time of one related to the other. The CCF is built with the cross-correlation coefficients, that is, the correlation coefficients between the elements in the two series analyzed, at different time displacements [25,26].

If we consider two time series \( z_t \) and \( x_t \), the cross-correlation coefficient for delay (time displacement or lag) \( L \), represented as \( r(L) \), is
The CCF is a graphical representation of the cross-correlation coefficients for delays in an interval $[-K, +K]$. By varying $L$ above and below zero, we can obtain an image of the possible interaction between the two series, identifying the time displacements at which this influence is higher. For example, if only one significant coefficient appears at $L = 0$, this means that the interaction between the two series is instantaneous. If a significant cross-correlation coefficient appears in $L > 0$, this means that changes in $x_t$ affect $z_t$ with a delay of $L$ periods. If a significant cross-correlation coefficient appears in $L < 0$, this means that changes in $z_t$ affect $x_t$ with a delay of $L$ periods.

### 2.2 Data sources

We used the public data provided by the Spanish Ministry of Health for our study, accessible daily via web [27] and tabulated in Ref. [28]. The dataset selected included six series, corresponding to the total number of infected people detected, the cumulative figures for recovered patients, the cumulative figures for deaths, the number of patients in ICUs, and the number of hospitalized patients. In the period under study, the criteria for gathering data were maintained constant.

In Fig. 35.1, we can appreciate the evolution of the different variables included in our dataset. The different curves show the cumulative values evolution. We can appreciate changes in the speed of growth, but it is better for recognize the evolution to use daily values instead of cumulative figures (Fig. 35.2).

The number of infected people was highly dependent on the number of tests performed and was therefore discarded for our main forecasting study. The most consistent series out of the given data was the cumulative death figures, and this became the main target of this study. The number of patients in ICUs was also studied, given its key relevance to show the stress placed on the Spanish healthcare system. The series with the number of recovered patients was studied in its relationship with the infected series, and only the number of hospitalized patients was not considered in our analysis.

Fig. 35.2 shows the evolution of daily values for infected people and death toll. In both curves, three periods are clearly defined:

- A first exponential growth period of the disease expansion in the first part of the pandemic evolution. This period cover until approximately March 26 or 27. As restrictions to mobility were applied on March 14 and considering the incubation lapse of one to two weeks, this first part can be assimilated with the free evolution of the disease.
- The second period starts approx. March 28 and last until about April 3. The curve, with a slower growing speed stabilizes and reaches the maximum. The effect of confinement measures starts to impact the pandemic evolution.
Chapter 35 • The growth of COVID-19 in Spain. A view based on time-series

FIGURE 35.1 Evolution of the pandemic in Spain. Cumulative values. From February 25 to April 7.

FIGURE 35.2 Evolution of the pandemic in Spain. Daily values. From February 25 to April 7.
- The third period, starting around April 4 means the beginning of the decreasing zone of the disease evolution, when the impact of Government actions finally changes the shape of the pandemic curve.

The shape of the first of these three periods, with nearly exponential growth, can be appreciated with more detail in Figs. 35.3 and 35.4.

Data are included in Annex A, at the end of this chapter. We have included not only the series used in this chapter but also the complete file for the period considered.

3. Analysis of the daily death toll

For the analysis, we considered the daily death figures. Though data about COVID-19 became available on February 20, the first deaths were recorded on March 4. Our analysis started on March 25, as a certain amount of data was required for significant analysis.

As literature indicates that epidemic processes frequently follow an exponential pattern in their growth phase (Chowell et al. [29,30]), we decided to analyze not the original data but to study their logarithms instead. The structure of the adjusted model and forecast was in line with epidemiological literature.

![FIGURE 35.3 Exponential pattern of evolution in daily Infected, before March 27.](image-url)
The model was re-evaluated with each new piece of data, starting from zero, that is, starting from the model identification. This is the reason for the changes in the model used, given that these were the best models for each day.

The model was consistently an ARIMA model with two differences, autoregressive and a MA order of 1 or 2, thus varying between ARIMA(2,2,1) and ARIMA(1,2,2), except on March 30 where an ARIMA(0,2,5) fitted slightly better than the usual models.

In the first part of the period, the main aim was to forecast the peak in the death curve.

The Spanish Government enforced restrictions on people’s movements on March 14 and given an incubation period of 7–14 days for the virus, the impact of these restrictions only became apparent one to two weeks later. Thus, the evolution before March 22 may be considered as the evolution pattern previous to restrictive measures.

The date for the maximum number of deaths and the forecast value for these deaths began to be brought forward as of March 25, 10 days after the confinement measures were brought in and when the death toll series has a length of at least 20 time periods (first dead was registered in March 4), a minimum to being able to perform a reasonably consistent analysis. On March 25, after an especially bad figure of persons death this day, the peak was expected on April 14 with a daily death toll of more than 5000. By the end of March, this date had been brought forward to April 3, with a death toll of about 1000. Finally, the maximum number of deaths occurred on April 2, with the terrible figure of
950 persons passing away in 1 day. These results are shown in Table 35.1, where, retrospectively, we have added forecasts for March 19 to March 24.

These values require some comment, given that the forecasting horizon to identify this peak is higher than the reasonable forecasting horizon. Predicting the evolution 10 days ahead (like on March 19) or even 24 days ahead (like on March 21) was more a mere mathematical projection than a reliable forecast. This was especially true for the first part of the series, as the quality of the forecast depends on the length of the time series. The identification of the curve’s peak after March 25, with horizons of less than 10 days, began to offer more reliable results, as the series became longer, and the forecast horizon grew shorter.

Once the confinement started to generate the desired effect, the model stabilized in an ARIMA(2,2,1)/ARIMA(1,2,2) for at least 10 days, before starting to change again in the downward part of the death toll curve.

The shape of the forecast evolution is shown in Fig. 35.5, corresponding to different dates between March 20 and April 7 (March 20, 22, 26, 28, and 31, and April 2, 4, and 7). First figures correspond to the period before the death curve peaked and the last ones after this maximum figure.

Centering the analysis on short term forecasts, those more reliable, we have reviewed the accuracy of three days ahead values. One day ahead forecasts always included the real value in the 95% confidence interval. In fact, the mean number of absolute sigma (as a distance between real data and forecasted value) was 0.35 and always less than 1. The same was valid for two days ahead forecasts (mean of absolute sigma between forecast and actual value of 0.32, always less than 1) and for three days ahead forecasts (mean of

Table 35.1 Forecast of daily deaths (* correspond to real values).

| Date of forecast | Model       | Maximum date predicted | Forecast | LCL  | UCL  |
|------------------|-------------|------------------------|----------|------|------|
| March 19, 2020   | ARIMA(1,2,2)| March 29, 2020         | 331      | 36   | 3,012|
| March 20, 2020   | ARIMA(1,2,2)| April 04, 2020         | 943      | 114  | 7,812|
| March 21, 2020   | ARIMA(1,2,2)| April 14, 2020         | 3,945    | 342  | 45,561|
| March 22, 2020   | ARIMA(2,2,1)| March 30, 2020         | 668      | 136  | 3,289|
| March 23, 2020   | ARIMA(2,2,1)| March 29, 2020         | 619      | 131  | 2,930|
| March 24, 2020   | ARIMA(1,2,2)| March 31, 2020         | 790      | 176  | 3,551|
| March 25, 2020   | ARIMA(2,2,1)| April 14, 2020         | 5,436    | 315  | 93,901|
| March 26, 2020   | ARIMA(1,2,2)| April 07, 2020         | 1,870    | 275  | 12,698|
| March 27, 2020   | ARIMA(2,2,1)| April 07, 2020         | 1,513    | 423  | 5,408|
| March 28, 2020   | ARIMA(1,2,2)| April 03, 2000         | 1,219    | 371  | 3,999|
| March 29, 2020   | ARIMA(1,2,2)| April 03, 2020         | 1,040    | 297  | 3,634|
| March 30, 2020   | ARIMA(0,2,5)| April 03, 2020         | 946      | 247  | 3,627|
| March 31, 2020   | ARIMA(1,2,2)| April 03, 2020         | 996      | 312  | 3,184|
| April 01, 2020   | ARIMA(2,2,1)| April 01, 2020*        | 864*     | —    | —    |
| April 02, 2020   | ARIMA(2,2,1)| April 02, 2020*        | 950*     | —    | —    |
FIGURE 35.5 Death toll evolution forecasted of March 20 (A), 22 (B), 26 (C), 28 (D), 31 (E), April 2 (F), 4 (G), and 7 (H).
Table 35.2  Three days ahead forecast with deviation from real value in number of sigma.

| Date of forecast | One day ahead forecast | Real value | σ   | Two days ahead forecast | Real value | σ   | Three days ahead forecast | Real value | σ   |
|------------------|------------------------|------------|-----|-------------------------|------------|-----|---------------------------|------------|-----|
| 19/03            | 143                    | 235        | −0.65 | 205                     | 324        | −0.58 | 207                      | 394       | −0.73 |
| 20/03            | 214                    | 324        | −0.59 | 305                     | 394        | −0.33 | 327                      | 462       | −0.42 |
| 21/03            | 320                    | 394        | −0.31 | 451                     | 462        | −0.03 | 511                      | 514       | −0.01 |
| 22/03            | 419                    | 462        | −0.16 | 514                     | 514        | 0.00  | 535                      | 738       | −0.43 |
| 23/03            | 491                    | 514        | −0.08 | 559                     | 738        | −0.43 | 576                      | 655       | −0.17 |
| 24/03            | 555                    | 738        | −0.46 | 628                     | 655        | −0.07 | 673                      | 769       | −0.19 |
| 25/03            | 804                    | 655        | 0.34  | 1029                    | 769        | 0.48  | 1225                     | 832       | 0.53  |
| 26/03            | 934                    | 769        | 0.33  | 965                     | 832        | 0.25  | 1147                     | 838       | 0.48  |
| 27/03            | 930                    | 832        | 0.21  | 939                     | 838        | 0.21  | 1105                     | 812       | 0.50  |
| 28/03            | 982                    | 838        | 0.30  | 1029                    | 812        | 0.38  | 1126                     | 849       | 0.39  |
| 29/03            | 985                    | 812        | 0.37  | 976                     | 849        | 0.26  | 1036                     | 864       | 0.31  |
| 30/03            | 938                    | 849        | 0.21  | 846                     | 864        | −0.04 | 894                      | 950       | −0.10 |
| 31/03            | 991                    | 864        | 0.27  | 949                     | 950        | 0.00  | 996                      | 932       | 0.11  |
| 01/04            | 816                    | 950        | −0.33 | 810                     | 932        | −0.30 | 757                      | 809       | −0.13 |
| 02/04            | 906                    | 932        | −0.06 | 940                     | 809        | 0.31  | 914                      | 674       | 0.54  |
| 03/04            | 868                    | 809        | 0.15  | 861                     | 674        | 0.51  | 784                      | 637       | 0.35  |
| 04/04            | 871                    | 674        | 0.54  | 747                     | 637        | 0.34  | 722                      | 743       | −0.05 |
| 05/04            | 736                    | 637        | 0.31  | 582                     | 743        | −0.52 | 562                      | 757       | −0.51 |
| 06/04            | 606                    | 743        | −0.48 | 520                     | 757        | −0.87 | 477                      | 683       | −0.72 |
| 07/04            | 518                    | 757        | −0.86 | 556                     | 683        | −0.47 | 453                      | 605       | −0.54 |

absolute sigma between forecast and actual value of 0.36, always less than 1). Table 35.2 shows three days ahead forecasts with the corresponding real values and the forecast error measured in sigma.

The expressions used to obtain these forecasts are the application in each case (in each daily model) of Eq. (35.3) or (35.4). Model parameters values changed daily, sometimes by the change in the ARIMA orders but also when orders maintain the same values, the incorporation of new data to the series produces different parameters’ estimations.

As an example, two daily forecasts functions will be presented, for March 20 and for April 2. In March 20 (see Table 35.1), selected model was an ARIMA(1,2,2), and thus the polynomial expression was:

\[(1 - \varphi_1 B)(1 - B)^2z_t = (1 - \theta_1 B - \theta_2 B^2)a_t\]

Once developed, the forecasting direct expression is:

\[z_t = (2 + \varphi_1)z_{t-1} - (1 + 2\varphi_1)z_{t-2} + \varphi_1 z_{t-3} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} + \text{Constant}\]

And using the parameter estimates (\(\varphi_1 = -0.58698; \ \theta_1 = 1.84775; \ \theta_2 = -0.88044\), constant = −0.02104 (not significant))
\[ z_t = 1.41302 z_{t-1} + 0.17396 z_{t-2} - 0.058698 z_{t-3} + a_t - 1.84775 a_{t-1} + 0.88044 a_{t-2} - 0.02104 \]

In April 2, the best model was an ARIMA(2,2,1), and the corresponding expressions are:

\[ (1 - \phi_1 B - \varphi_1 B^2)(1 - B)^2 z_t = (1 - \theta_1 B)a_t \]

\[ z_t = (2 + \varphi_1)z_{t-1} - (1 + 2\varphi_1 - \varphi_2)z_{t-2} + (\varphi_1 - 2\varphi_2)z_{t-3} + \varphi_2 z_{t-4} + a_t - \theta_1 a_{t-1} + \text{Constant} \]

And taking into account the parameter estimates (\( \varphi_1 = -1.11585; \varphi_2 = -0.473501; \theta_1 = 1.00713, \) constant = \(-0.0382876\) (not significant))

\[ z_t = 0.88415 z_{t-1} - 2.705201 z_{t-2} + 0.168848 z_{t-3} - 0.473501 z_{t-4} + a_t - 1.00713 a_{t-1} - 0.0382876 \]

4. Analysis of the relationship between deaths and intensive care unit figures

Patients with more serious problems and more complicated prognoses were taken into the ICUs, which became the real bottleneck, saturating hospitals in Spain, and in all countries [31,32]. There is obviously a link between the series of deaths and that of ICU admissions.

ARIMA analysis brings us a tool to verify this relationship and to understand how ICU figures affect death toll values. To do this, we analyzed the CCF between the ARIMA residuals from the death toll adjustment and the ICU admissions series. The CCF enables us to detect the possible delay of the impact of one series on the other. In our case, the cause-effect relationship is from ICUs to deaths, and in the CCF computation, we should detect important values for some cross-correlation coefficients in negative delays.

Fig. 35.6 shows the CCF obtained on April 7. We selected the last day of the period under study to perform our analysis so as to take advantage of the maximum amount of data, but in previous dates, the shape was similar.

As Fig. 35.6 shows, there is a clear maximum in CC coefficients for delay \(-8\). Symmetric coefficients around delay \(-8\) are called usually “satellites” and are an analytical effect of the basic relation in the main delay.

The meaning of this pattern has an important interpretation: ICU admissions had an impact on the number of deaths with a delay of eight days. More precisely, the part of the evolution of the death toll values that cannot be explained by its own history, is, at least in part, explained by the number of ICU admissions eight days before.

5. Relationship between infected and recovered

There should be a relationship between the figures of infected people and those of recovered people. Causal models as SIR or SEIR formulates this relationship in their defining expressions [33]. We will evaluate this using CCF from the ARIMA perspective. If
we compare cumulative curves, the relationship will show a high correlation coefficient, but this can be a spurious value (when computing the relationship between any two growing series, the result will be a high-correlation coefficient). Nevertheless, by logical and medical reason, a causal relation should exist and requires some analysis.

We have used the series of daily new infected persons and daily recovered persons. CCF has been used to evaluate how both series interact. To do this, first an automated forecasting model selection has been made for the evolution of daily recovered series, resulting in a weak structure known as random walk, or in terms of ARIMA models an ARIMA(0,1,0) model. CCF has been obtained then for the residuals of this model vs the series of daily infected. We have used the same time window, and the result shows that today’s recovered values have a significative dependence on values of infected cases with delays 2 and 8 (with a 95% level of confidence). Fig. 35.7 shows this CCF.
6. Conclusions and final comments

The period considered for this study is especially relevant because it corresponds to the period in which the Spanish healthcare system moved from a situation of normality to one of maximum stress, on the verge of saturation and sometimes even beyond this boundary. After the peak in the pandemic curve, the healthcare system still had a lot of work, but the stress started to decrease, and the situation began to return to normal.

As we have seen, ARIMA models were able to adjust the evolution pattern of the COVID-19 growth phase for the Spanish case. Models with usually no more than three parameters were capable of generating accurate forecasts with a mean error of about 0.33 sigma, for one, two and three days ahead forecasts.

The use of the CCF enabled us to detect an interesting relationship between the number of ICU admissions and the number of deaths eight days later, identifying maybe a critical phase in the time spent in ICUs.

Also using CCF has showed that there is a significant relationship between recovered persons and infected cases with 2 and 8 days of delay, maybe linked with less severe cases and medium sever cases, as for severe cases the length of stay in hospital is higher, ever more than one month. These severe cases are not detected by the CCF because the length of the series involved in the analysis.

The analysis has continued after the period under study in this paper. The situation has changed in the downward phase of the pandemic’s evolution in Spain. On one hand, there have been certain issues with data gathering, as data are recorded by the regional governments and then sent to the central government, and this process has suffered some discrepancies. In addition, the structure of the model has changed and will be reviewed in later research papers.

Annex A. Data

| Date          | Cumulative infected | Cum. Recovered | Cum. Deceases | Cum. ICU | Cum. Hospitalized |
|---------------|---------------------|----------------|---------------|----------|-------------------|
| February 25, 2020 | 3                   |                |               |          |                   |
| February 26, 2020 | 10                  |                |               |          |                   |
| February 27, 2020 | 16                  |                |               |          |                   |
| February 28, 2020 | 32                  |                |               |          |                   |
| February 29, 2020 | 44                  |                |               |          |                   |
| March 01, 2020   | 66                  |                |               |          |                   |
| March 02, 2020   | 114                 |                |               |          |                   |
| March 03, 2020   | 135                 |                |               |          |                   |
| March 04, 2020   | 198                 | 1              | 7             |          |                   |
| March 05, 2020   | 237                 | 3              | 9             |          |                   |
| March 06, 2020   | 365                 | 5              | 11            |          |                   |
| March 07, 2020   | 430                 | 8              |              |          |                   |
| March 08, 2020   | 589                 | 17             |              |          |                   |
| March 09, 2020   | 999                 | 23             | 17            | 68       |                   |
March 10, 2020 1,622 135 35 100
March 11, 2020 2,128 183 47 142
March 12, 2020 2,950 189 84 190
March 13, 2020 4,209 189 120 272
March 14, 2020 5,753 517 136 293
March 15, 2020 7,753 517 84 190
March 16, 2020 9,191 530 309 432
March 17, 2020 11,178 1,028 491 563 3,215
March 18, 2020 13,716 1,081 598 774 5,717
March 19, 2020 17,147 1,107 767 939 5,717
March 20, 2020 19,980 1,585 1,002 1,141 10,542
March 21, 2020 24,926 2,125 1,326 1,612 10,542
March 22, 2020 28,572 2,575 1,720 1,785 15,554
March 23, 2020 33,089 3,355 2,182 2,355 18,273
March 24, 2020 39,673 3,794 2,696 2,636 22,762
March 25, 2020 47,610 5,367 3,434 3,166 26,960
March 26, 2020 56,188 7,015 4,089 3,679 31,912
March 27, 2020 64,059 9,357 4,858 4,165 36,293
March 28, 2020 72,248 12,285 5,690 4,575 40,630
March 29, 2020 78,797 14,709 6,528 4,907 43,397
March 30, 2020 85,195 16,780 7,340 5,231 46,617
March 31, 2020 94,417 19,259 8,189 5,607 49,243
April 01, 2020 102,136 22,647 9,053 5,872 51,418
April 02, 2020 110,238 26,743 10,003 6,092 54,113
April 03, 2020 117,710 30,513 10,935 6,416 56,637
April 04, 2020 124,736 34,219 11,744 6,532 57,612
April 05, 2020 130,759 38,080 12,418 6,861 58,744
April 06, 2020 135,032 40,437 13,055 6,931 59,662
April 07, 2020 140,510 43,208 13,798 7,069 63,093

References

[1] N. Zhu, D. Zhang, W. Wang, W. Wang, X. Li, B. Yang, J. Song, et al., A novel coronavirus from patients with pneumonia in China, N. Engl. J. Med. 382 (February 20, 2020) 727–733.
[2] D. Wu, T. Wu, Q. Liu, Z. Yang, The SARS-CoV-2 outbreak: what we know, Int. J. Infect. Dis. 94 (May 2020) 44–48.
[3] M. Xie, Q. Chen, Insight into 2019 novel coronavirus — an updated interim review and lessons from SARS-CoV and MERS-CoV, Int. J. Infect. Dis. 94 (May 2020) 119–124.
[4] B. Robson, Computers and viral diseases. Preliminary bioinformatics studies on the design of a synthetic vaccine and a preventative peptidomimetic antagonist against the SARS-CoV-2 (2019-nCoV, COVID-19) coronavirus, Comput. Biol. Med. 119 (April 2020). Article 103670.
[5] T.L. Le, et al., The COVID-19 vaccine development landscape, Nat. Rev. Drug Discov. 19 (May 2020) 305–306.
[6] C. Ma, S. Su, J. Wang, L. Wei, S. Jiang, From SARS-CoV to SARS-CoV-2: safety and broad-spectrum are important for coronavirus vaccine development. Microb. Infect. 22 (July–August 2020) 245–253.
Chapter 35 • The growth of COVID-19 in Spain. A view based on time-series

[7] D.C. Hall, H.F. Ji, A search for medications to treat COVID-19 via in silico molecular docking models of the SARS-CoV-2 spike glycoprotein and 3CL protease. Trav. Med. Infect. Dis. 35 (May–June 2020) 101646.

[8] M.P. Lythgoe, P. Middleton, Ongoing clinical trials for the management of the COVID-19 pandemic. Trends Pharmacol. Sci. 41 (No. 6) (June 2020) 363–382.

[9] J. Gao, X. Tian, X. Yang, Breakthrough: chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies, BioSci. Trends 14 (No. 1) (March 16, 2020) 72–73.

[10] M. Cascella, M. Rajnik, A. Cuomo, et al., Features, Evaluation and Treatment Coronavirus (COVID-19) [Accesed April 12, 2020]. in: StatPearls [Internet], StatPearls Publishing, Treasure Island (FL), January 2020. Available from: https://www.ncbi.nlm.nih.gov/books/NBK554776/.

[11] C. Calvo, M. García, J.C. Carlos, J.L. Vázquez, Recommendations on the clinical management of the COVID-19 infection by the «new coronavirus» SARS-CoV2. Anales de Pediatría (English Edition) 92 (No. 4) (April 2020) 241.e1–241.e11.

[12] D.L. McKee, A. Sternberg, U. Stange, S. Laufer, C. Naujokat, Candidate drugs against SARS-CoV-2 and COVID-19, Pharmacol. Res. 157 (July 2020). Article 104859.

[13] A. Briz-Redón, A. Serrano-Aroca, A spatio-temporal analysis for exploring the effect of temperature on COVID-19 early evolution in Spain, Sci. Total Environ. 728 (August 2020). Article 138811.

[14] A.S. Ahmar, E. Boj, SutteARIMA: short-term forecasting method, a case: covid-19 and stock market in Spain, Sci. Total Environ. 729 (August 2020). Article 138883.

[15] J. Yuan, M. Li, G. Lv, Z.K. Lu, Monitoring transmissibility and mortality of COVID-19 in Europe, Int. J. Infect. Dis. 95 (June 2020) 311–315.

[16] D. Benvenuto, M. Giovanetti, L. Vassallo, S. Angeletti, M. Ciccozzi, Application of the ARIMA model on the COVID-2019 epidemic dataset, Data in Brief 29 (April 2020) 105340.

[17] H. Tandon, P. Ranjan, T. Chakraborty, V. Suhag. Coronavirus (COVID-19): ARIMA Based Time-Series Analysis to Forecast Near Future. arXiv:2004.07859 [q-bio.PE]. Submitted on April 16, 2020. (Accessed May 2020).

[18] A. Kumar, P.K. Gupta, A. Srivastava, A review of modern technologies for tackling COVID-19 pandemic, Diabetes & Metab. Syndr. Clin. Res. Rev. 14 (Issue 4) (July–August 2020) 569–573.

[19] J.S. Armstrong, Evaluating forecasting methods, in: J.S. Armstrong (Ed.), Principles of Forecasting. International Series in Operations Research & Management Science, vol. 30, Springer, Boston, MA, 2001.

[20] M. Joffe, M. Gambhir, M. Chadeau-Hyam, P. Vineis, Causal diagrams in systems epidemiology. Emerg. Themes Epidemiol. 9 (March 19, 2012) 1.

[21] E.B. Postnikov, Estimation of COVID-19 dynamics “on a back-of-envelope”: does the simplest SIR model provide quantitative parameters and predictions? Chaos Solit. Fractals 135 (June 2020). Article 109841.

[22] G.E.P. Box, G.W. Jenkins, G.C. Reinsel, G. Ljung, Time Series Analysis: Forecasting and Control, fifth ed., John Wiley & Sons, 2016, p. 712.

[23] P.J. Brockwell, R.A. Davis, Introduction to Time Series and Forecasting, third ed., Springer Texts in Statistics, Switzerland, 2016, p. 428.

[24] J.D. Cryer, K.S. Chan, Time Series Analysis with Applications in R, second ed., Springer Texts in Statistics, New York, 2016, p. 501.

[25] L.D. Haugh, Checking the independence of two covariance-stationary time series: a univariate residual cross-correlation approach, J. Am. Stat. Assoc. 71 (354) (1976) 378–385.
[26] K. Bhaskaran, A. Gasparini, S. Hajat, L. Smeeth, B. Armstrong, Time series regression studies in environmental epidemiology, Int. J. Epidemiol. 42 (4) (August 2013) 1187–1195.

[27] I. de Salud Carlos III, M. de Sanidad de España (n.d.). https://covid19.isciii.es/. Accessed during March and April 2020.

[28] Github. Datasets in, n.d. https://github.com/datadista/datasets/tree/master/COVID%2019 Accessed during March and April 2020.

[29] G. Chowell, C. Viboud, J.M. Hyman, L. Simonsen, The western Africa ebola virus disease epidemic exhibits both global exponential and local polynomial growth rates, PLOS Currents Outbreaks (January 2015) 1–25. https://doi.org/10.1371/currents.outbreaks.8b55f4bad99ac5c5db3663e916803261.

[30] G. Chowell, L. Sattenspiel, S. Bansal, C. Viboud, Mathematical models to characterize early epidemic growth: a review, Phys. Life Rev. 18 (September 2016) 66–97.

[31] ECDC (European Centre for Disease Prevention and Control), Coronavirus Disease 2019 (COVID-19) Pandemic: Increased Transmission in the EU/EEA and the UK — Seventh Update, March 25, 2020 (Stockholm).

[32] J. Phua, L. Weng, L. Ling, M. Egi, C.M. Lim, J.V. Divatia, et al., Intensive care management of coronavirus disease 2019 (COVID-19): challenges and recommendations, Lancet Respir. Med. 8 (5) (May 2020) 506–518.

[33] A. Naheed, M. Singh, D. Lucy, Numerical study of SARS epidemic model with the inclusion of diffusion in the system, Appl. Math. Comput. 229 (February 25, 2014) 480–498.