Available online at http://scik.org

J. Math. Comput. Sci. 2022, 12:10

https://doi.org/10.28919/jmcs/6541

ISSN: 1927-5307

THE APPLICATION OF ARIMA MODEL TO ANALYZE COVID-19 INCIDENCE PATTERN IN SEVERAL COUNTRIES

JACQUES DEMONGEOT[1,*], KAYODE OSHINUBI[1], MUSTAPHA RACHDI[1], LAHOUCINE HOBBAD[2], MOHAMED ALAHIANE[2], SIHAM IGGUI[1], JEAN GAUDART[3], IDIR OUASSOU[2]

[1]Laboratoire AGEIS EA 7407, Faculté de Médecine, Université Joseph Fourier, Grenoble, La Tronche, France

[2]Laboratoire de Modélisation des Systèmes Complexes (LMSC), Université Cadi Ayyad, ENSA, Marrakech, Maroc

[3]SESSTIM UMR 1252, Faculté de Médecine, Université Aix-Marseille, Marseille, France

Abstract: The COVID-19 pandemic continues to spread and already shows a recurrence in many countries, despite several social distancing and vaccination measures implemented all around the world. Epidemiological data are available, and we use the Auto-Regressive Integrated Moving Average (ARIMA) model to analyze incidence pattern and to generate short-term forecasts of cumulative reported cases in Morocco, France, Italy, Spain and USA, using daily reported cumulative cases data from Worldometers, and we report 5-day and 10-day ahead forecasts of cumulative cases and check a posteriori the precision of this forecasting, by confronting it to the real data observed. In the discussion, we propose a link between the ARIMA, elevation and average temperature in several countries' modelling approaches, for allowing the comparison between their explicative abilities.

Keywords: coronavirus; epidemic; forecasts; cumulative cases; ARIMA model; COVID-19.

2010 AMS Subject Classification: 00A71, 93D05.

*Corresponding author

E-mail address: Jacques.Demongeot@univ-grenoble-alpes.fr

Received July 22, 2021

1. INTRODUCTION

The novel coronavirus outbreak (COVID-19), as named by the World Health Organization (WHO) on 11th February 2020, began in Hubei Province, China, in December 2019 and continues to cause infections in multiple countries. The COVID-19 represents the newest zoonotic Coronavirus disease that crossed species to affect humans and spread in an unprecedented manner.

The outbreak was declared a pandemic and a Public Health Emergency on 30 January 2020, by WHO. To control this pandemic, the governments have enacted a range of social distancing strategies, such as city-wide lockdowns and isolation of suspected cases plus a vaccination policy. The numbers of cumulated cases and deaths continue to accumulate every day and despite a slowing due to strict lockdowns combined with isolation, quarantine measures and vaccination, many countries entered in second, third until fourth waves of the outbreak Seligmann et al. [38]. While the transmission potential of this novel coronavirus can reach high values, the epidemiological features as dependencies on geo-climatic or demographic variables, and the mechanisms of transmission and host susceptibility of the viral agent SARS-CoV-2 of COVID-19 outbreak are still unclear (Demongeot et al., [10]; Demongeot and Seligmann [11]; Seligmann et al., [38]). In this paper, we use the Auto-Regressive Integrated Moving Average (ARIMA) model to generate 5-day and 10-day ahead forecasts of the cumulative reported cases in the countries as Morocco, France, Italy, Spain, UK and USA.

Currently, several mathematical methods are applied in disease incidence prediction such as linear regression, artificial neural networks and grey box models. The ARIMA model is commonly used in infectious disease time series prediction, especially for series that have a cyclic or repeated pattern. The model was conceived for economics applications, but is well convenient in the medical field nowadays. The principle of the model contains filtering out the high-frequency noise in the data, detecting local trends based on linear dependence and forecasting by extrapolating trends as it has been already done for some countries in the COVID-19 case, in order to explain for example correlations observed with geo-climatic parameters as mean temperature and elevation (Deb and Majumdar [1]; Demongeot et al., [10-13]; Perez et al., [14]; Faye et al., [15]; Ilie et al., [24]; Behambar et al., [30]; Seligmann et al., [38]). Despite its high predictive performance, the model has some limitations which decrease its scope of application. The model assumes a linear relationship between the dependent and independent variables while the actual data often present non-linear relationships. Besides, the model assumes that the mean and variance of time series are independent of time, which means stationary of order 2. Thus, more than one approach should be tested to choose the better one as in former studies for some epidemics (Kane et al., [25]; Luo et al., [28]; Rubaihayo et al., [35]; Soebiyanto et al., [40]; Wei et al., [42]; Abioye et al., [45]).

We present in Section 2 Methods the sources of data used in the present paper and the model proposed. Then, in Section 3 Results, we provide the main observations coming from the ARIMA approach and in Section 4 Discussion, we discuss the role of age in the classical Ordinary Differential Equation (ODE) modelling of infectious diseases (Demongeot et al., [4-9]; Gaudart et al., [16-19]; Liu et al., [27], Scarpino and Petri [36]), for making explicit a common theoretical basis. Eventually, in Section 5 Conclusions, we give some perspectives to this work, in terms of long-term monitoring of the COVID-19 pandemic, in particular with regard to its temporo-spatial diffusion (Griette et al., [20]; Guttmann et al., [21-23]).

2. MATERIALS AND METHODS

2.1 Empirical analysis

2.1.1 Data sources

The daily incidence data of COVID-19 from January to May 2020 were collected from the European Centre for Disease Prevention and Control [43] and Worldometers [44] sites.

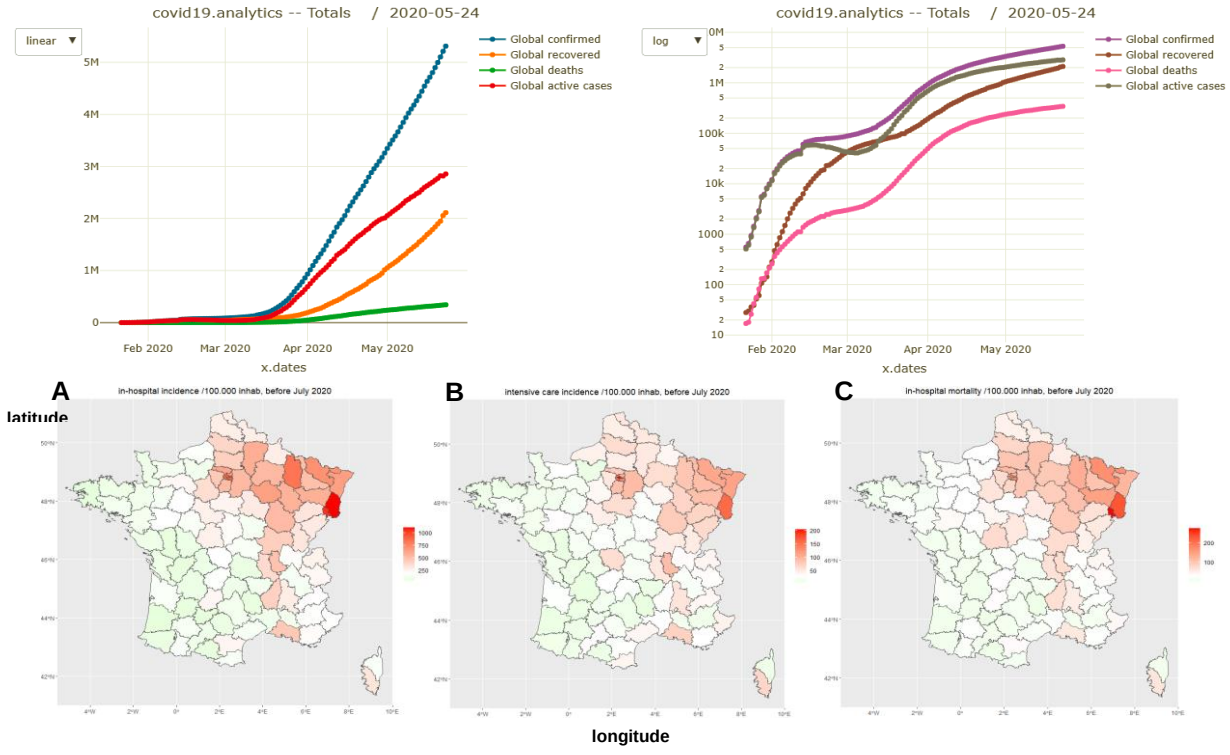


FIGURE 1: *Top left*: COVID-19 world global cumulated cases (confirmed, recovered, active and deaths) with linear scale. *Top right*: same data as on the left with logarithmic scale. *Bottom*: before July 2020, in France, rates/100,000 for A: in-hospital incidence, B: intensive care incidence, C: in-hospital mortality.

Data from January 1 to May 10 2020 were used to build the ARIMA model, data from May 11 to May 16 2020 to evaluate the forecasting precision by the model, and before July for showing the geoclimatic dependencies of the SARS-CoV-2 incidence (Figure 1). Simulations are using Gnu-R ARIMA software.

2.1.2 Study and visual representation of data

When predicting the evolution of a time series process, it is useful to find a model generating past values in order to extrapolate the simulation results of this model to the future. For performing that task, the model must sufficiently describe the past, and be robust in its predictions. The time series we aim to predict is the number of cumulated confirmed cases $Y_{i,k}$ of the COVID-19 outbreak at day i in country or region k. The time unit is the day. A first approach to time series data confirms that the logarithmic transformation allows us to describe the phenomenon in a functionally simple manner and with the added advantage of expressing the evolution of the number of cases in logarithmic form.

The series $\log(Y_{i,k})$ appears on Figure 1 Top right at first glance to suggest a non-linear evolution close to the quadratic one. It should be pointed out that this quadratic evolution would only describe the phenomenon at the start of epidemy, and would no longer be the generating model, given that $Y_{i,k}$ is following a saturation dynamics. The adequate model is a lifecycle model where the maximum is generated in a finite time and not in an infinite time like in the classic logistic model.

The ARMA models contain auto-regressive moving average model (ARMA), auto-regressive integrated moving average model (ARIMA) and seasonal autoregressive integrated moving average model (SARIMA). The most sophisticated model is the SARIMA S(p,d,q; P,D,Q,s) model, where p means the order of auto-regression, d the degree of trend difference, q the order of moving average, P the seasonal auto-regression lag, D the degree of seasonal difference, Q the seasonal moving average, and s the length of the cyclical pattern Wei et al., [42]. In this paper, we use only the auto-regressive integrated moving average ARIMA(p,d,q) method with time-dependent parameters. Time series stationarity, parameter estimation, model checking and prediction will be done to establish this ARIMA model as in (Demongeot et al., [10]; Luo et al., [28]; Rubaihayo et al., [35]), by using like in [10] the statistical software facility given in: www.statsmodels.org/stable/generated/statsmodels.tsa.arima_model.ARIMA.html.

The goal of the study is to identify if the coefficient estimates for the time trend change after a certain point, and if so, to examine it further to find out the nature of change and potential causes, with the following model:

$$\log(\Theta_{i,k}) = f(i,\tau_k,k) + \sum_{j=1,3} \gamma_j L_{j,i,k} + U_{i,k} \quad (1)$$

where $\Theta_{i,k} = (Y_{i,k}-Y_{i-1,k})/P_{i,k}$ denotes the incidence rate of Covid-19 of region k at time i in {1,…,T} (every time point i denotes a single day), $P_{i,k}$ denotes the overall population of region k at time i, $f(i,\tau_\kappa,k)$ is the trend function and τ_k is the day when the first confirmed case is observed in the region k, i.e., we have: $Y_{i,k} = 0$ for $i < \tau_k$ and $Y_{i,k} \geq 0$ for $i \geq \tau_k$. $L_{j,i,k}$ is the dummy variable signifying the lockdown, γ_j captures the effect of lockdown, for j = 1,2,3 and $U_{i,k}$ is the error process. In the real data, it is observed that the number of cumulated cases grows with time and then stabilizes after a certain time. Since we consider the following structure for $\log(\Theta_{i,k})$ in the model (1), by assuming an ARIMA(p,d,q) structure for the error process, while f is supposed to be a polynomial quadratic trend in time $(i-\tau_\kappa)$. In particular, the coefficients for the linear and quadratic terms in f are considered to be different for $i-\tau_k<\eta_k$ and $i-\tau_k\geq\eta_k$.

The parameter η_k is estimated from the data and it tells us when the trend of the growth changes its pattern in region k. Combining everything, we will use the following equation for the model:

$$\log(\Theta_{i,k}) = ß_0 + ß_{1,i,k}(i-\tau_k) + ß_{2,i,k}(i-\tau_k)^2 + \Sigma_{j=1,3}\,\gamma_j\,L_{j,i,k} + \Sigma_{m=1,p}\,a_m\log(\Theta_{i-m,k}) + \Sigma_{m=1,q}\,\gamma_m U_{i-m,k} + U_{i,k} \quad (2)$$

Here, p and q denote respectively the auto-regressive and moving-average orders of the ARIMA error process and $U_{i,k} = \varepsilon_{i,k}$ denotes a standard normal white noise process, and we have:

- for n=1,2, $ß_{n,i,k} = ß_{n,1,k}$, and for $i<\eta_k$, $ß_{n,i,k} = ß_{n,2,k}$ for $i\geq\eta_k$ (3)
- for j=1,2,3, $L_{j,i,k}=0$ for $i<\zeta_{j,k}$, and $L_{j,i,k}=1$, for $i\geq\zeta_{j,k}$, (4)

where ζ is a binary indicator of pre $(\zeta_{j,k}=0)$ and post $(\zeta_{j,k}=1)$ infection waves j=1,2,3. We estimate η_k, $\zeta_{j,k}$ and $\varepsilon_{j,k}$ $(1<j<3)$, p and q from data using Akaike information criterion (AIC).

The best ARIMA model for each country was studied in the paper as parameters p=6, d=1 and q=0 as in Demongeot et al., [10].

3. RESULTS

3.1 Descriptive statistics

3.1.1 The basic reproduction number R_0

There are several methods for estimating the basic reproduction number R_0, equal to the mean number of new infections caused by an infected person at day j among the susceptible population. The time-dependent method Obadia et al., [31] has been used for estimating R_0 at time t, denoted R(t), which allows to the detection of the end of the first wave, e.g., the 18th of May 2020 in Morocco (Figure 2).

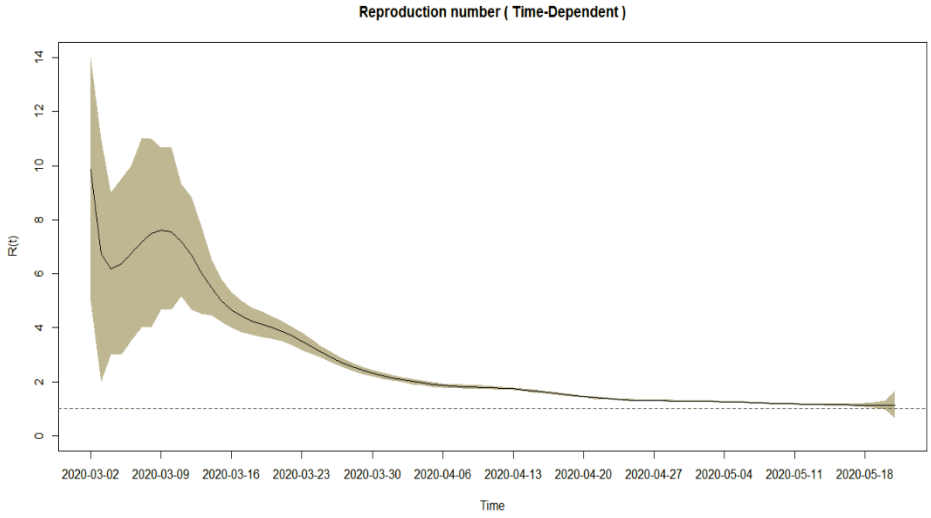


FIGURE 2: Time-dependent estimation method of the parameter R_0 at time t, denoted R(t), along the first wave of the Covid-19 outbreak in Morocco. The zone in grey corresponds to the 95%-confidence set of the R(t)'s.

3.1.2 The data for 6 countries

The dynamics of the COVID-19 outbreak can be daily followed in all world countries from European Centre for Disease Prevention and Control [43] and Worldometers [44] web sites. On Figure 3, we can see comparable evolution of the first wave in 4 European countries and in Morocco, showing comparable trends along February/March 2020, period of the first wave for all these countries.

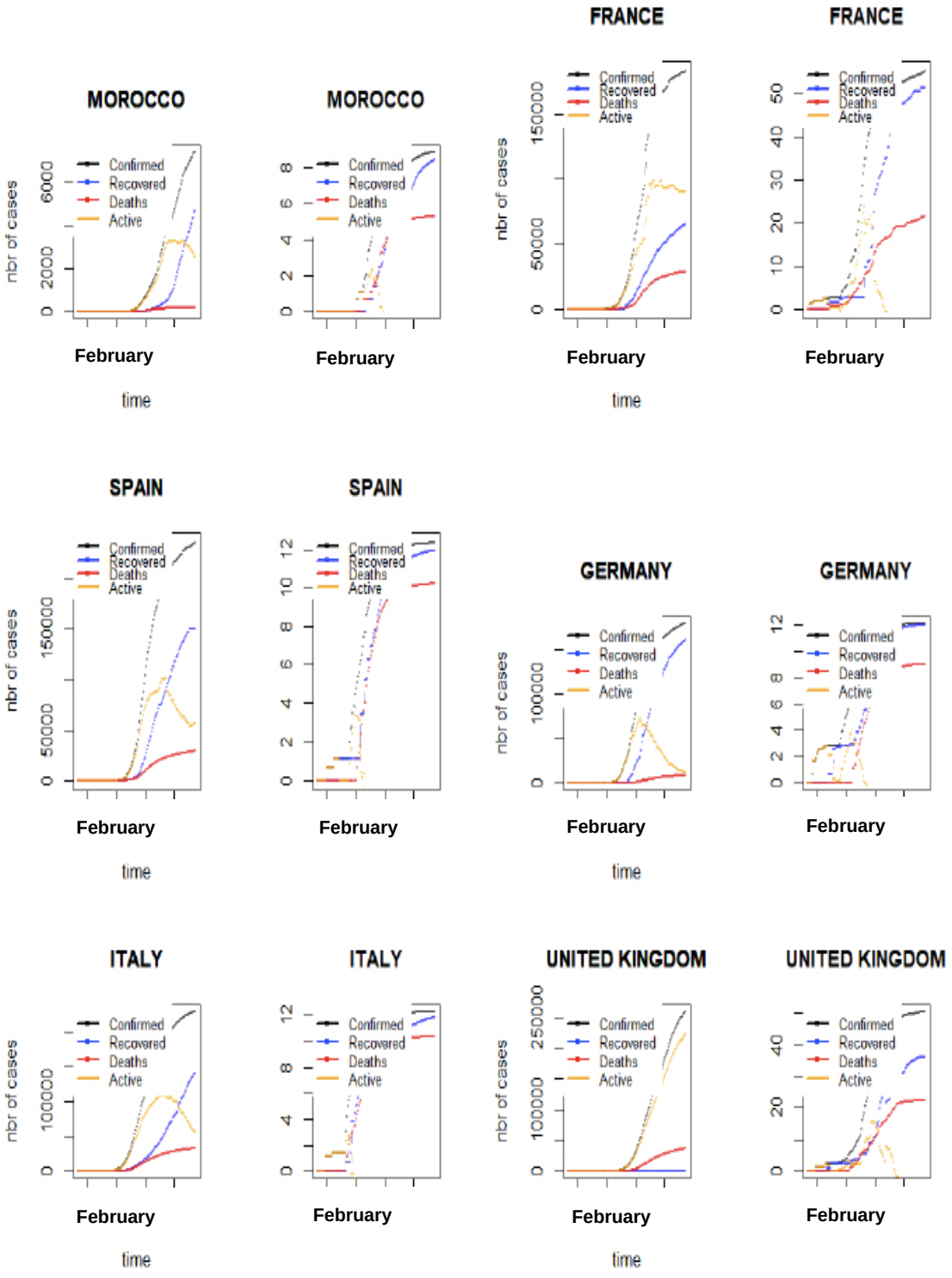


FIGURE 3: Dynamic evolution of the COVID-19 outbreak in 5 European countries, France, Spain, Germany, Italy, United Kingdom compared to a North-African country, Morocco. Left graphs correspond for each country to cumulated cases numbers and right graphs to their logarithm.

3.2 ARIMA model for the first wave

By using the data concerning the new daily cases of the COVID-19 outbreak coming from the Worldometers web site [44], it is possible to study their correlation with geoclimatic variables like the mean temperature or the mean elevation, and demographic variables like the median age in many countries (Demongeot et al., [10]; Seligmann et al., [38]). For that purpose, the ARIMA technique allows for i) extracting the trend using the moving average method with a long window (Figure 4 top), then study the stationarity of the series obtained by differentiation with this trend (Figure 4 bottom) and eventually iii) show if the residue can be considered as a Gaussian noise with zero mean or if there is still a seasonal component to be extracted by a moving average of window length adapted to the value of its period.

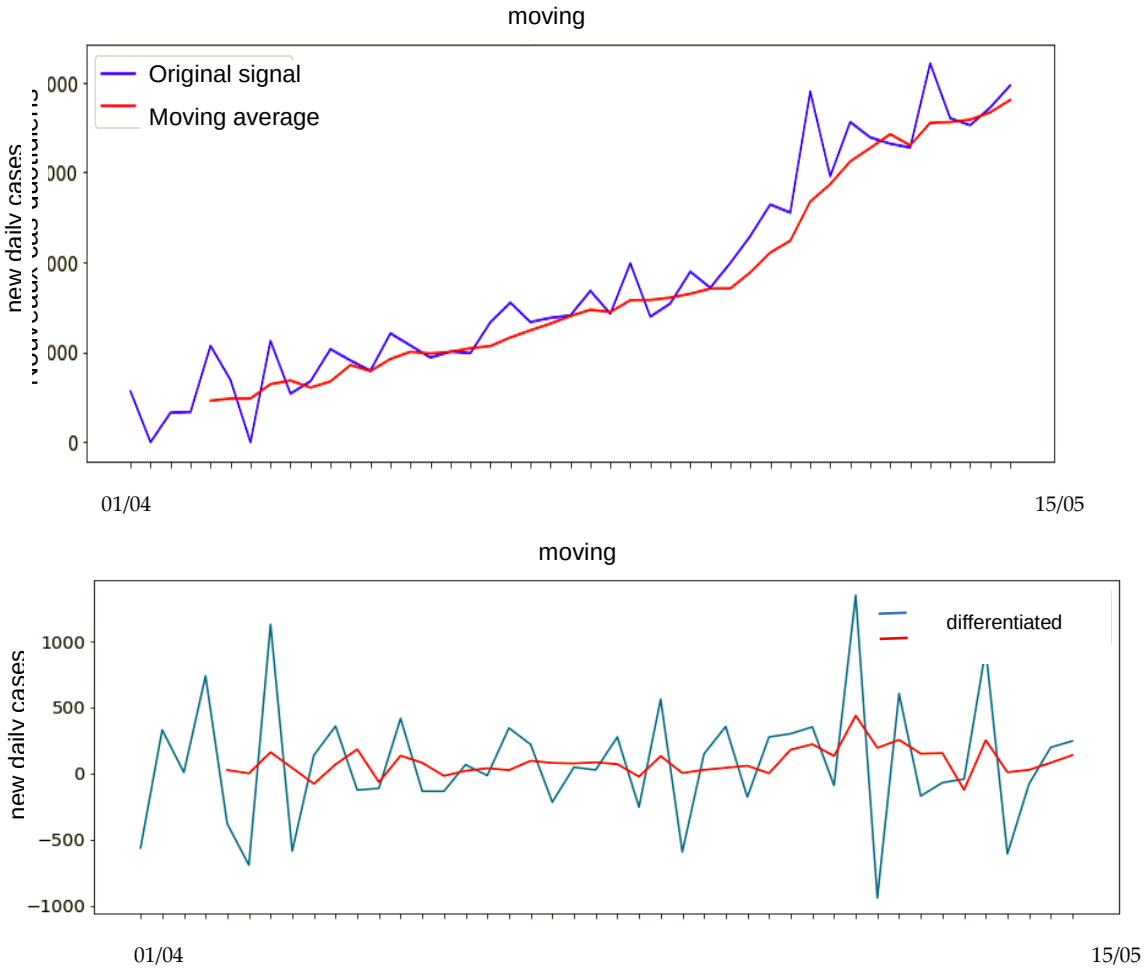


FIGURE 4: Top: New daily cases of the COVID-19 (in blue) with indication of the trend (in red) calculated by using the moving average method. Bottom: same series obtained by subtracting the trend (in blue) and indication of the moving average (in red).

The first wave of the COVID-19 outbreak occurred at the same time on the beginning of February 2020 in France and on the beginning of April 2020 for many other countries like Qatar and India (Figure 6). The ARIMA software gives the auto-correlation function of the original new cases (Figure 6), and the estimation of its initial slope used for studying correlations (Table 1 and Figure 7), showing for elevation less than 1000 m a decrease of the auto-decorrelation length for new cases (estimated by the initial slope of the auto-correlation function), corresponding to a diminution of the contagiousness period (Demongeot et al., [11]; Seligmann et al., [39]).

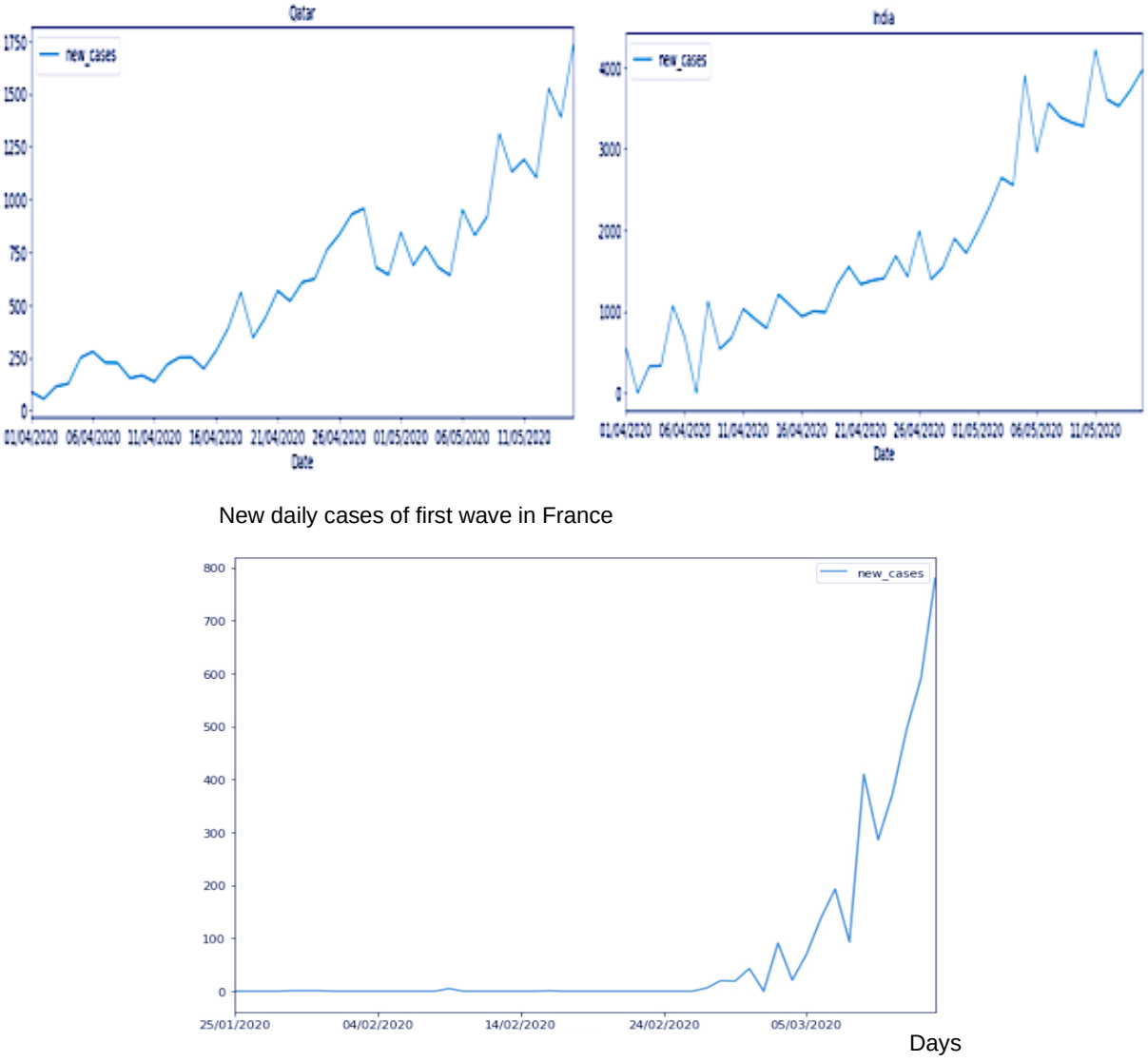


FIGURE 5: Top: New daily cases of the COVID-19 first wave in Qatar (left) and India (right). Bottom: same in France.

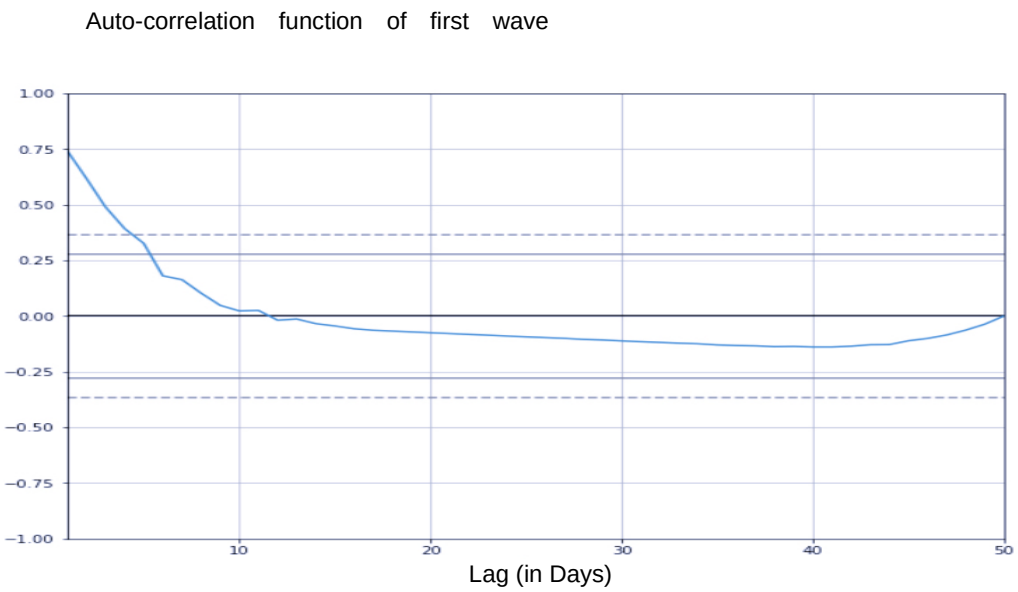


FIGURE 6: Auto-correlation function of the COVID-19 first wave in France.

Table 1: Initial slope of the autocorrelation function of the ARIMA model, and mean temperature in May 2020 (in °F), for 13 countries in first wave.

Country	Auto-correlation initial slope	Mean Temperature (°F)
Singapore	- 0.030	83.95
Saudi Arabia	- 0.060	81.89
Qatar	- 0.070	80.64
United Arab Emirates	- 0.080	80.12
Peru	- 0.010	78.68
India	- 0.050	77.75
Mexico	- 0.100	69.07
Pakistan	- 0.060	66.25
Columbia	- 0.040	59.27
Portugal	- 0.140	57.40
South Africa	- 0.110	56.92
France	- 0.110	52.11
Croatia	- 0.069	46.94
Ukraine	- 0.130	44.25

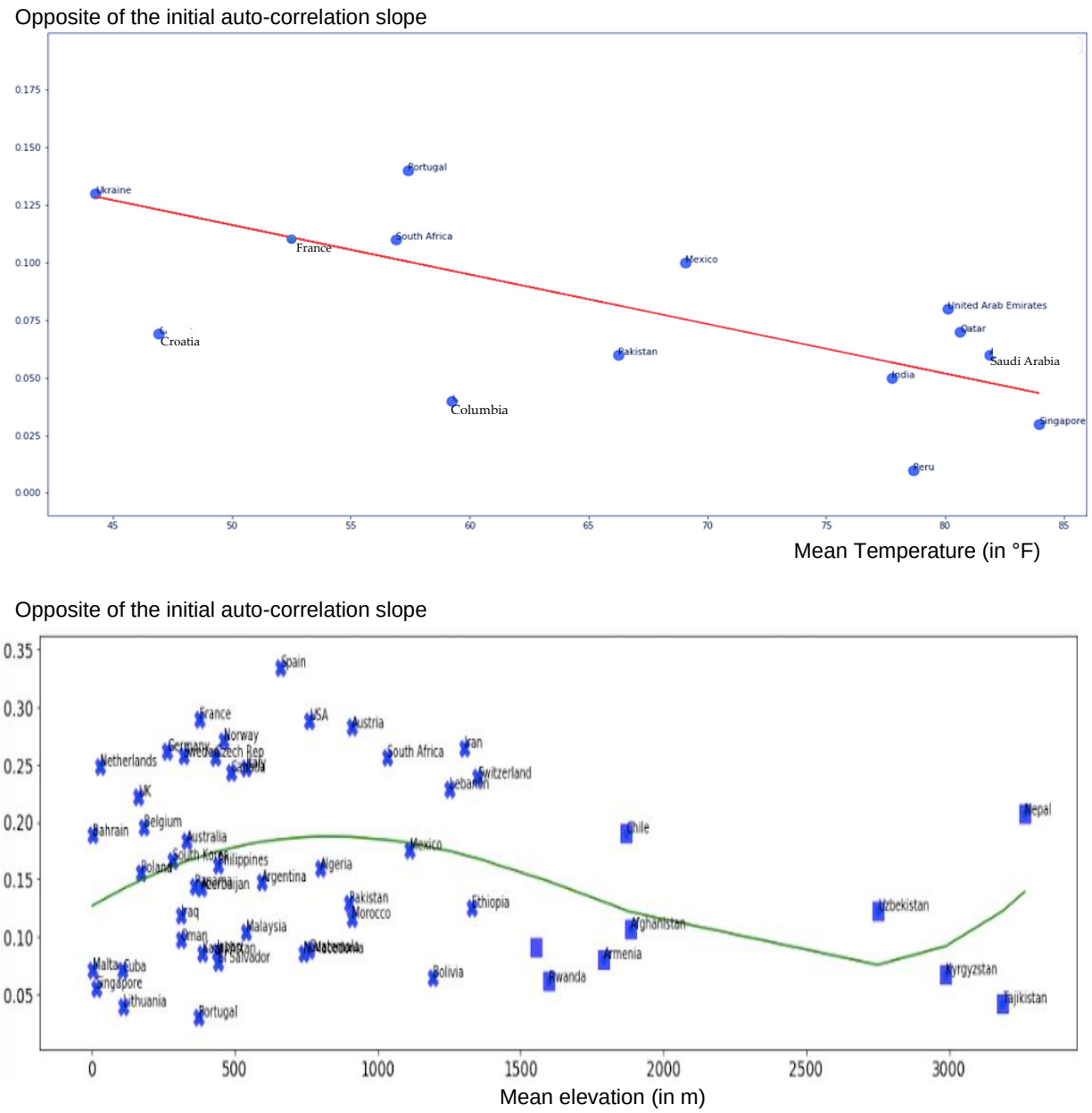


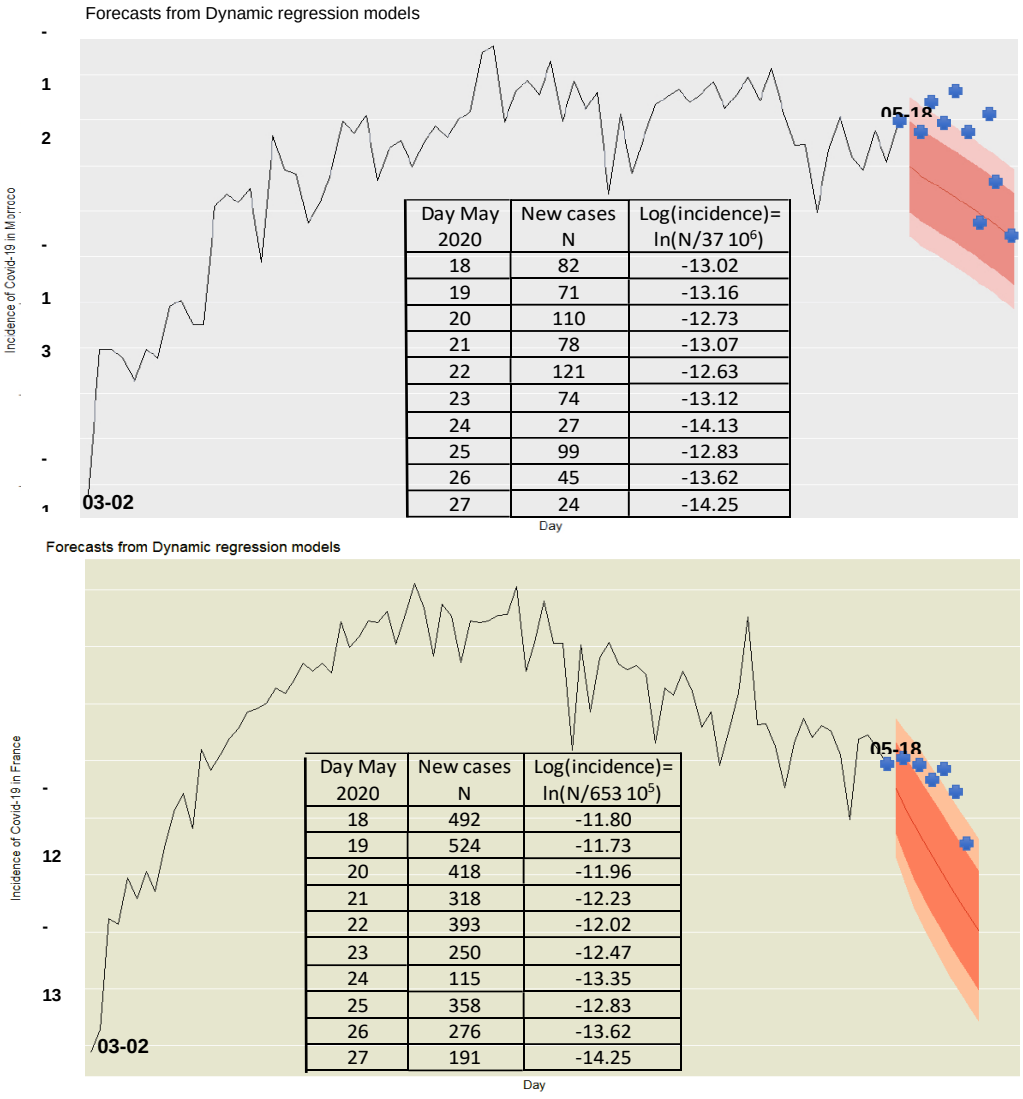
FIGURE 7: Top: Linear regression of the opposite of the initial slope of the first wave correlation function vs mean temperature, with correlation coefficient equal to R = - 0.7. Bottom: polynomial regression with mean elevation.

The opposite of the initial slope of the autocorrelation function A(j) (which measures the correlation between observations in a time series separated by j days) is greater the shorter the period of contagiousness, and therefore the lower the regressivity of the ARIMA model, this may be due to a decrease in the virulence of the virus, by alteration during its passage from the

transmitter (the patient in a period of contagiousness) to the recipient (the susceptible individual), in an atmosphere at high mean temperature and altitude, which destroys by heat and radiation the essential components of the virus (capsid and RNA). Regarding the altitude, the cubic adjustment shows a paradoxical effect at low altitude due to countries with high mean temperature and low altitude.

3.3. Forecasts for Morocco and France

It is possible to use the ARIMA model to do one-week forecasts of new COVID-19 cases. We have done it at the end of the first wave in Morocco and France (Figure 8), showing roughly the start of the decrease, but over-estimating the intensity of this trend.



Day May 2020	New cases N	Log(incidence)= ln(N/37 10^6)
18	82	-13.02
19	71	-13.16
20	110	-12.73
21	78	-13.07
22	121	-12.63
23	74	-13.12
24	27	-14.13
25	99	-12.83
26	45	-13.62
27	24	-14.25

Day May 2020	New cases N	Log(incidence)= ln(N/653 10^5)
18	492	-11.80
19	524	-11.73
20	418	-11.96
21	318	-12.23
22	393	-12.02
23	250	-12.47
24	115	-13.35
25	358	-12.83
26	276	-13.62
27	191	-14.25

FIGURE 8: Top: one week forecast of Log(incidence) in Morocco at the end of first wave. Pink regions correspond to the 90%- (dark pink) and 95%- (clear pink) confidence set for the predicted new cases. Bottom: same results for France.

In both cases, the forecast for Morocco and France shows an overestimation of the decrease at the end of the first wave possibly due to the influence of the periodic drop in counts of new daily cases at the end of the week, visible in many countries Demongeot et al., [12].

3.4 Second wave in Armenia

We can do the same study for the second wave of the COVID-19 outbreak as for the first one. On Figure 9, the example of Armenia shows a linear moving average (in red) canceling the periodic effect of partial settlements of weekends already observed for Morocco, to be subtracted from the original new cases (in blue) to obtain a stationary residue (in black). On Figure 10, the auto-correlation function of new cases is used for estimating the length of the contagiousness period, here about 8 days (at intersection with the value of 0.25, limit of the significance with p-value = 0.05 for a time series of 45 days).

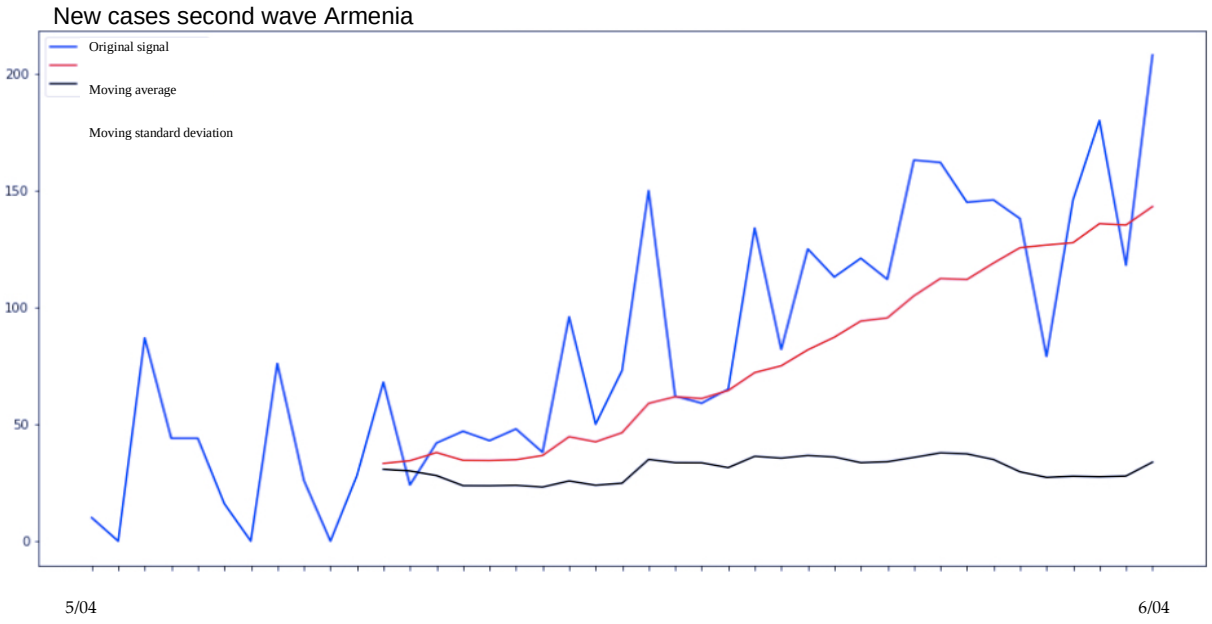


FIGURE 9: New cases of the COVID-19 second wave in Armenia.

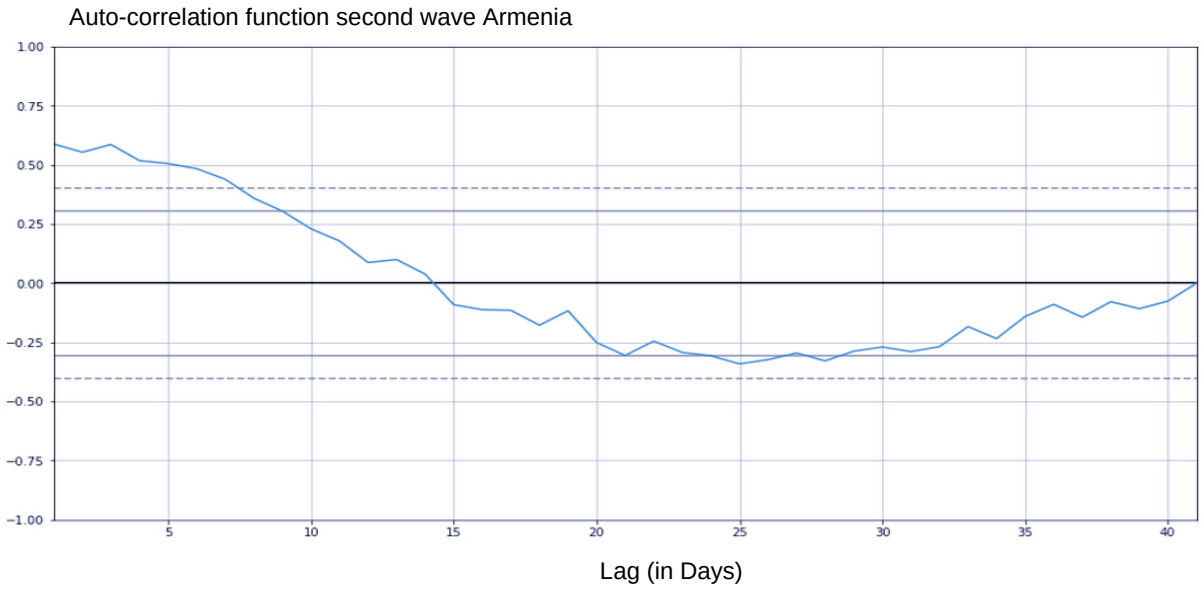


FIGURE 10: Auto-correlation function of the COVID-19 second wave in Armenia.

The study for different countries of the value of its initial slope (Table 2) shows an increase until 70°F, then a decrease with mean temperature more than 70°F, which corresponds to an adaptation of the viral pathogenicity caused by numerous mutations observed at high altitude due for example to the mutagenic power of UVs (Figure 11). Many mutations with high contagiousness have in fact been observed in countries with a high mean temperature (Brazil, Colombia, South Africa).

Table 2: Initial slope of the autocorrelation function of the ARIMA model, and mean temperature (in °F), for 21 countries in second wave.

Country	Autocorrelation curve slope (averaged on 4 first days)	Mean Temperature (°F)
Armenia	- 0.090	44.850
Lithuania	- 0.230	46.860
Czech Rep.	- 0.197	51.000
South Korea	- 0.090	54.000
Chile	- 0.090	56.428
Portugal	- 0.190	56.570
Argentina	- 0.240	57.210
Algeria	- 0.100	57.710
Kenya	- 0.310	63.800
Azerbaijan	- 0.130	65.860
Kazakhstan	- 0.210	66.000
Macedonia	- 0.230	66.280

Malaysia	- 0.260	68.570
Iraq	- 0.150	70.570
Uzbekistan	- 0.170	74.500
Malta	- 0.330	74.920
Lebanon	- 0.180	75.420
Kyrgyzstan	- 0.200	76.290
Iran	- 0.140	82.850
Sri Lanka	- 0.160	84.860
Oman	- 0.130	92.210

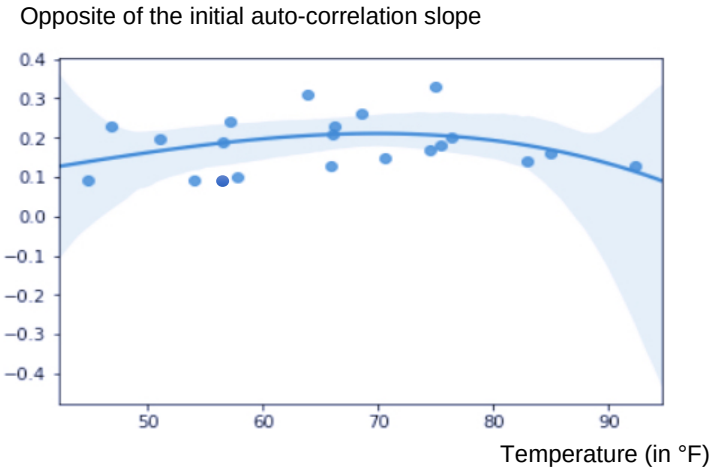


FIGURE 11: Parabolic regression of the opposite of the initial slope vs temperature. The blue zone corresponds to the 95%-confidence set in the neighborhood of the regression parabola.

4. DISCUSSION

4.1 Modelling the influence of age

The role of the age on virulence is clearly proved with inverted effects observed in the second wave (Demongeot et al., [10], Liu et al., [26]; Scapino and Petri [36]; Statista [37]) and this influence can be modelled. If we consider only two age classes (young and old), the SIR equations Demongeot et al., [4-6] will become:

$$\begin{aligned} dS_1/dt &= -ß_{11}S_1I_1 - ß_{12}S_1I_2 + k_1R_1 - kS_1 + fS_1 - \mu S_1 \\ dI_1/dt &= ß_{11}S_1I_1 + ß_{12}S_1I_2 - (1-\mu_1)I_1 - \mu_1 I_1 \\ dR_1/dt &= (1-\mu_1)I_1 - k_1R_1 \\ dS_2/dt &= -ß_{21}S_2I_1 - ß_{22}S_2I_2 + k_2R_2 + kS_1 - \mu S_2 \\ dI_2/dt &= ß_{21}S_2I_1 + ß_{22}S_2I_2 - (1-\mu_2)I_2 - \mu_2 I_2 \end{aligned} \quad (5)$$

$$dR_2/dt = (1-\mu_2)I_2 - k_2R_2.$$

where S_1 represents the age class of young and adults strictly less than 60 years, whose size is 48 million in France [37] and S_2 represents the age class of more than 60 years, whose size is about three times less, that is 17.3 million [37]. The SIR equations link the sizes of susceptible (S), infected (I) and recovered (R) populations. If we will assume that the entry (by birth f) or exit (by natural death μ) rates are offset over the duration where the epidemic wave studied (f=μ), then the SIR equations satisfy the conservation equation:

$$d(S_1 + I_1 + S_2 + I_2 + R_1 + R_2)/dt = fS_2 - \mu S_2 - \mu_1 I_1 - \mu_2 I_2 = 0,$$

only if the mortality due to the infection is neglectable: $\mu_1 = \mu_2 = 0$. If not, we can have a diminution of the initial total population size at the asymptotic stationary state, depending on the rate of transmissions ($ß_{ik}$), loss of immunity rates (k_i) and aging parameter (k). An example is shown on Figure 13, where the initial sizes of the susceptible classes respect the ratio between young and adults (S_1) and elderly (S_2), the values of the SIR model parameters being chosen by assuming the absence of mitigation measures: we suppose that every day a person of S_1 establishes 24 contacts ($c_{1,1}$ = 18 contacts with persons from S_1 and $c_{1,2}$ = 6 contacts with persons from S_2) and a person of S_2 has 15 contacts ($c_{2,1}$ = 12 contacts with persons from S_2 and $c_{2,2}$ = 3 contacts with persons from S_1). We choose f=3, μ=0.1, $ß_{1i} = 0.03c_{1,i}$ and $ß_{2i} = 0.02c_{1,i}$ for i=1,2, and unrealistic no-immunization (k_1 = 0.8, k_2 = 0.5) and death rates (μ_1 = 0.6, μ_2 = 0.9) rates, for showing the consequences of an important fatality epidemic without mitigation, vaccination and/or therapy (Figure 12).

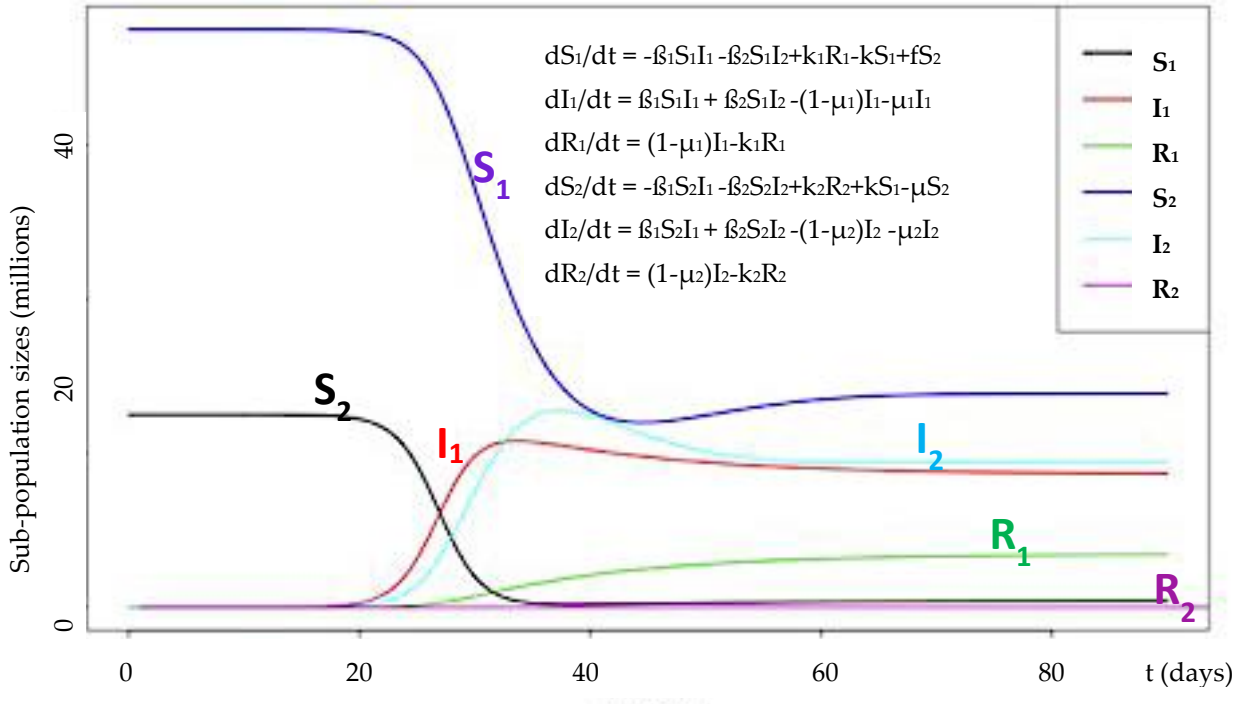


FIGURE 12: SIR model with 2 age classes.

4.2 Median age and incidence of COVID-19

The dependence of the incidence of COVID-19 on median age in the observed countries can change between two consecutive waves as shown in Figures 13 and 14, probably due to an adaptation of the virus to the demographic profile of the population in which it propagates, with a positive correlation of the incidence vs the median age (R=0.41) during the first wave (Figure 13) and a negative correlation (R = - 0.41) during the second wave (Figure 14). One possible explanation lies in i) the exhaustion of the targets most sensitive to viruses (elderly people with chronic comorbidities) and ii) the application of mitigation measures (prevention, protection, eviction) between the two waves, which mainly concerned classes at risk, including the elderly.

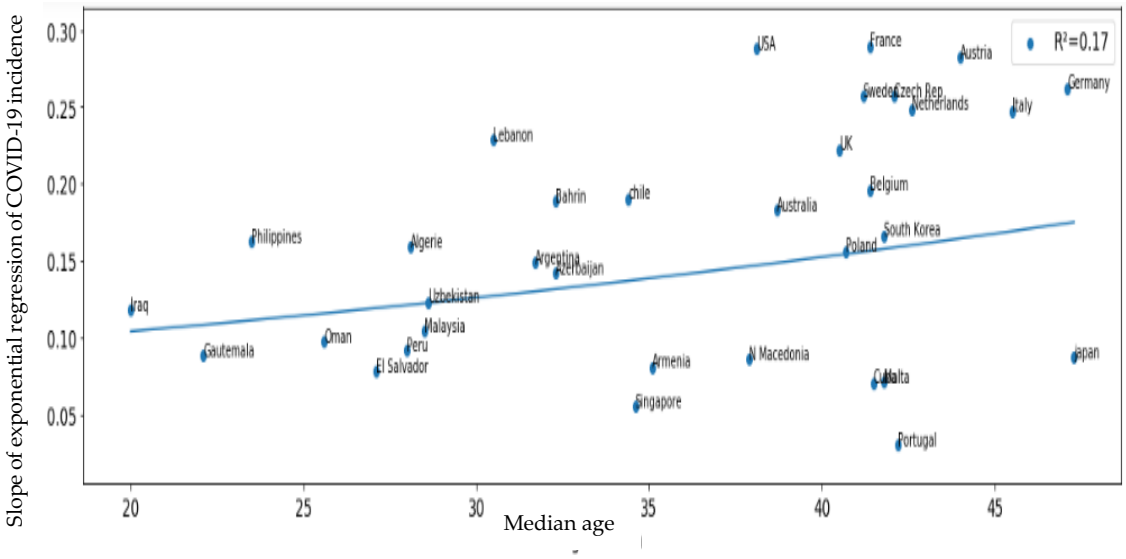


FIGURE 13: Positive correlation (R=0.41) between the slope of the exponential regression of the COVID-19 prevalence and the median age of the countries observed during the first wave.

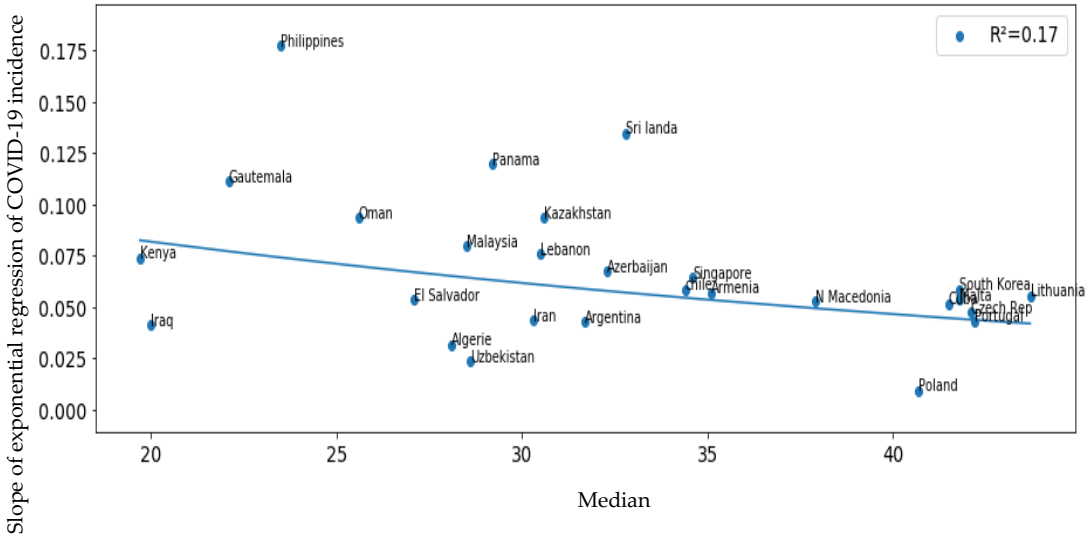


FIGURE 14: Negative correlation (R=-0.41) between the slope of the exponential regression of the COVID-19 prevalence and the median age of the countries observed during the second wave.

5. CONCLUSION

The third wave of the COVID-19 outbreak started already in many countries and forecasting its start has not been possible (like for the start of the second wave) from the ARIMA approach, but the data about new cases are already showing similarities with the first wave, in particular for those concerning the correlation with temperature and elevation in the countries in which it occurred Seligmann et al., [38-39].

The same work has to be made for the third and fourth waves as for the two first ones, notably in the directions discussed in Section 4, in what concerns the influence of four factors on the outbreak dynamics: i) the age, because an adaptation has been observed during the second wave which seems to concern younger patients leading to discuss the value of R_0 in a heterogeneous population, ii) the duration of the contagiousness, which seems to be longer than in previous waves, iii) the entropy of the distribution of the daily reproduction rates (as already pointed out in Demongeot et al., [12]; Oshinubi et al., [32-33]; Rhodes and Demetrius [34]) and which may correspond to a change in the immune defense sequence, with suppression of the apparent improvement due to innate immunity (causing a U-shaped profile of the distribution, hence a decrease of its entropy) and iv) the influence of geoclimatic but also of socio-economic factors. Eventually, the frequent over-deviation of the new daily cases observed in third and fourth waves in many countries (pos-

sibly due to the entanglement of several successive or simultaneous health measures like distancing, lockdown and vaccination) would lead to replace in the future ARIMA models by generalized additive models (GAM) with a negative binomial regression. Indeed, descriptive ARIMA models and are known not to replace explanatory models based on plausible contagion mechanisms, such as ODE models and finding the best model to represent the COVID-19 data remains an open challenge despite notable advances in this direction (Demongeot et al., [13]; Griette et al., [20]; Wei et al., [42]). The spatial diffusion of SARS Cov-2 and its variants is also an important subject, not addressed here. Only Figure 1 Bottom shows a spatial difference between the French regions. To address this problem, we can, on a descriptive level, make use of the statistical techniques of spatial interpolation by kriging already used for malaria in (Gaudart et al., [16-17]) and of detection of spatial heterogeneities in public health data Guttmann et al., [21-23]. On the explanatory level, the spatiotemporal modeling using the partial differential equations (PDE) (Gaudart et al., [17-18]) would make it possible to estimate the speed of propagation, as well as its direction (NorthEast / SouthWest for the first wave in France). The analysis of the spatio-temporal heterogeneities mentioned above would also make it possible to propose a model containing delays between the causes (like date of contagion) and the effects (like limits of the period of the subsequent contagiousness), and also to work on the structure of the noise, cause of intrinsic fluctuations in epidemic data (in particular linked to variations in daily reproduction rates Demongeot et al., [12]), which is only addressed here through the residue of the ARIMA model. All of these important points will be covered in future articles.

APPENDIX

Table 3: Description of the parameters

Parameter	Description
k	index representing a country or a region
i	index representing a time-point (every i denotes a single day)
R_0	number of cases directly caused by one patient suffering COVID-19
I	time interval between infection and subsequent transmission
τ_k	day when the first confirmed case is observed in region k
γ_j	parameter capturing the effect of lockdown for infection waves j=1,2,3
η_k	parameter indicating when trend changes its pattern in region k

Variable	Description
$X_{i,k}$	number of confirmed new cases in region k at time i
$D_{i,k}$	number of deaths in region k at time i
$R_{i,k}$	number of people recovered in region k at time i
$P_{i,k}$	population size (in millions) of region k at time i
$\zeta_{j,k}$	binary indicator of pre $(\zeta_{j,k}=0)$ and post $(\zeta_{j,k}=1)$ infection waves j=1,2,3
$L_{j,i,k}$	dummy variable signifying the lockdown at time i in region k for wave j
$P_{i,k} = Y_{i,k}/P_{i,k}$	proportion (prevalence) of total confirmed cases of region k at time i
$\Theta_{i,k} = (Y_{i,k}-Y_{i-1,k})/P_{i,k}$	proportion of new cases (incidence) of region k at time i
$f(i,\tau,k)$	trend function, at time i, for first day τ and region k
$U_{i,k}$	stochastic error process at time i for region k
$\varepsilon_{j,k}$	standard normal white noise process for wave j and region k

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interest.

REFERENCES

[1] S. Deb, M. Majumdar, A time series method to analyze incidence pattern and estimate reproduction number of COVID-19, ArXiv:2003.10655 [q-Bio, Stat]. (2020).

[2] M. Delbrück, Statistical fluctuations in autocatalytic reactions, J. Chem. Phys. 8 (1940), 120-124.

[3] L. Demetrius, Boltzmann, Darwin and the directionality theory, Phys. Rep. 530 (2013), 1-86.

[4] J. Demongeot, J. Gaudart, A. Lontos, E. Promayon, J. Mintsa, M. Rachdi, Least diffusion zones in morphogenesis and epidemiology, Int. J. Bifurcation Chaos, 22 (2012a), 1250028.

[5] J. Demongeot, J. Gaudart, J. Mintsa, M. Rachdi, Demography in epidemics modelling, Commun. Pure Appl. Anal. 11 (2012b), 61-82.

[6] J. Demongeot, O. Hansen, A.S. Jannot, C. Taramasco, Random modelling of contagious (social and infectious) diseases: examples of obesity and HIV and perspectives using social networks, In: IEEE AINA'12, IEEE Proceedings, Piscataway, (2012c), 1153-1160.

[7] J. Demongeot, O. Hansen, H. Hessami, A.S. Jannot, J. Mintsa, M. Rachdi, C. Taramasco, Random modelling of contagious diseases, Acta Biotheor. 61 (2013a), 141-172.

[8] J. Demongeot, M. Ghassani, M. Rachdi, I. Ouassou, C. Taramasco, Archimedean Copula and Contagion Modeling in Epidemiology, Networks Heterogeneous Media, 8 (2013b), 149-170.

[9] J. Demongeot, M. Ghassani, H. Hazgui, M. Rachdi, The copule approach in epidemiologic modelling, In: ICPS'13, Springer, New York, (2015), 151-161.

[10] J. Demongeot, Y. Flet-Berliac, H. Seligmann, Temperature decreases spread parameters of the new covid-19 cases dynamics, Biology (Basel), 9 (2020), 94.

[11] J. Demongeot, H. Seligmann, Covid-19 and miRNA-like inhibition power, Med. Hypotheses, 144C (2020), 110245.

[12] J. Demongeot, K. Oshinubi, M. Rachdi, H. Seligmann, F. Thuderoz, J. Waku, Estimation of Daily Reproduction Rates in COVID-19 Outbreak, Computation, 9 (2021), 109.

[13] J. Demongeot, Q. Griette, P. Magal, Computations of the transmission rates in SI epidemic model applied to COVID-19 data in mainland China, R. Soc. Open Sci. 7 (2020), 201878.

[14] F.J. Diaz Perez, D. Chinarro, R. Pino Otin, R. Díaz Martín, M. Moises Diaz, A. Guardiola Mouhaffel, Comparison of Growth Patterns of COVID-19 Cases through the ARIMA and Gompertz Models, Case Studies: Austria, Switzerland, and Israel, Rambam Maimonides Med. J. 11 (2020), e0022.

[15] C. Faye, C.T. Wade, I.D. Dione, A dissymmetry in the figures related to the covid-19 pandemic in the world: what factors explain the difference between Africa and the rest of the world? J. Ongoing Chem. Res. 5 (2020), 61-70.

[16] J. Gaudart, O. Touré, N. Dessay,A.L. Dicko, D. Ranque, L. Forest, J. Demongeot, O.K. Doumbo, Modelling malaria incidence with environmental dependency in a locality of Sudanese savannah area, Mali, Malaria J. 8 (2009), 61.

[17] J. Gaudart, M. Ghassani, J. Mintsa, M. Rachdi, J. Waku, J. Demongeot, Demography and Diffusion in epidemics: Malaria and Black Death spread, Acta Biotheor. 58 (2010a), 277-305.

[18] J. Gaudart, M. Ghassani, J. Mintsa, J. Waku, M. Rachdi, O.K. Doumbo, J. Demongeot, Demographic and spatial factors as causes of an epidemic spread, the copule approach. Application to the retro-prediction of the Black Death epidemy of 1346, In: IEEE AINA' 10, IEEE Proceedings, Piscataway, (2010b), 751-758.

[19] J. Gaudart, J. Landier, L. Huiart, E. Legendre, L. Lehot, M.K. Bendiane, L. Chiche, A. PetitJean, E. Mosnier, F. Kirakoya-Samadoulougou, J. Demongeot, R. Piarroux, S. Rebaudet, Factors associated with spatial heterogeneity of Covid-19 in France: a nationwide ecological study, The Lancet Public Health 6 (2021), e222-e231.

[20] Q. Griette, J. Demongeot, P. Magal, A robust phenomenological approach to investigate COVID-19 data for France, Math. Appl. Sci. Eng. 2 (2021), 14031.

[21] A. Guttmann, L. Ouchchane, X. Li, I. Perthus, J. Gaudart, J. Demongeot, J.Y. Boire, Performance map of a cluster detection test using the extended power, Int. J. Health Geographics, 12 (2013), 47.

[22] A. Guttmann, X. Li, J. Gaudart, Y. Gérard, J. Demongeot, J.Y. Boire, L. Ouchchane, Spatial heterogeneity of type I error for local cluster detection tests, Int. J. Health Geographics, 13 (2014), 15.

[23] A. Guttmann, X. Li, F. Feschet, J. Gaudart, J. Demongeot, J.Y. Boire, L. Ouchchane, Cluster detection tests in spatial epidemiology: a global indicator for performance assessment, PLoS ONE, 10 **(**2015), e0130594.

[24] O.D. Ilie, R.O. Cojocariu, A. Alin Ciobica, S.I. Timofte, Mavroudis, I.; Bogdan Doroftei, B. Forecasting the Spreading of COVID-19 across Nine Countries from Europe, Asia, and the American Continents Using the ARIMA Models, Microorganisms, 8 (2020), 1158.

[25] M.J. Kane, N. Price, M. Scotch, P. Rabinowitz, Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks, BMC Bioinform. 15 (2014), 276.

[26] K. Liu, Y. Chen, R. Lin, K. Han, Clinical features of COVID-19 in elderly patients: A comparison with young and middle-aged patients, J. Infection, 80 (2020), e14-e18.

[27] Z. Liu, P. Magal, O. Seydi, G. Webb, Understanding Unreported Cases in the COVID-19 Epidemic Outbreak in Wuhan, China, and the Importance of Major Public Health Interventions, Biology (Basel), 9 (2020), 50.

[28] L. Luo, L. Luo, X. Zhang, X. He, Hospital daily outpatient visits forecasting using a combinatorial model based on ARIMA and SES models, BMC Health Services Res. 17 (2017), 469.

[29] J. Mintsa, M. Rachdi, J. Demongeot, Stochastic approach in modelling epidemic spread, In: IEEE AINA' 11, IEEE Proceedings, Piscataway, (2011), 478-482.

[30] H. Noghani Behambari, M. Salari, F. Noghani, N. Tavassoli, The Effect of Temperature on COVID-19 Confirmed Cases: Evidence from US Counties (June 27, 2020), SSRN, (2020), 3636662.

[31] T. Obadia, R. Haneef, P.Y. Boëlle, The R_0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. BMC Med. Inform. Decision Making, 12 (2012), 147.

[32] K. Oshinubi, M. Rachdi, J. Demongeot, Analysis of daily reproduction rates of COVID-19 using Current Health Expenditure as Gross Domestic Product percentage (CHE/GDP) across countries, Healthcare, 9 (2021), 1247.

[33] K. Oshinubi, F. Al-Awadhi, M. Rachdi, J. Demongeot, Data Analysis and Forecasting of COVID-19 Pandemic in Kuwait, MedRxiv, (2021), 21261059.

[34] C.J. Rhodes, L. Demetrius, Evolutionary Entropy Determines Invasion Success in Emergent Epidemics, PLoS ONE, 5 (2010), e12951.

[35] J. Rubaihayo, N.M. Tumwesigye, J. Konde-Lule, F. Makumbi, Forecast analysis of any opportunistic infection among HIV positive individuals on antiretroviral therapy in Uganda, BMC Public Health, 16 (2016), 766.

[36] S.V. Scarpino, G. Petri, On the predictability of infectious disease outbreaks, Nature Commun. 10 (2019), 898.

[37] Statista. Available online: https://fr.statista.com/statistiques/472349/repartition-population-groupe-dage-france/ (accessed on 8th September 2020), (2020).

[38] H. Seligmann, S. Iggui, M. Rachdi, N. Vuillerme, J. Demongeot, Inverted covariate effects for mutated 2nd vs 1st wave Covid-19: high temperature spread biased for young, Biology (Basel), 9 (2020), 226.

[39] H. Seligmann, N. Vuillerme, J. Demongeot, Unpredictable, Counter-Intuitive Geoclimatic and Demographic Correlations of COVID-19 Spread Rates, Biology, 10 (2021), 623.

[40] R.P. Soebiyanto, F. Adimi, R.K. Kiang, Modeling and predicting seasonal influenza transmission in warm regions using climatological parameters, PLoS One, 5 (2010), e9450.

[41] J. Wallinga, P. Teunis, Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures, Am. J. Epidemiol. 160 (2004), 509.

[42] W.D. Wei, J.J. Jiang, H. Liang, L. Gao, B.Y. Liang, J.G. Huang, N. Zang, Y. Liao, J. Yu, J. Lai, F. Qin, J. Su, L. Ye, H. Chen, Application of a combined model with autoregressive integrated moving average (ARIMA) and generalized regression neural network (GRNN) in forecasting hepatitis incidence in Heng County, China, PloS One, 11 (2016), e156768.

[43] www.ECDC. Available online: https://www.ecdc.europa.eu/en (accessed on 8th April 2021), (2021).

[44] www.Worldometers. Available online: https://www.worldometers.info/coronavirus/ (accessed on 8th April 2021), (2021).

[45] A.I. Abioye, M.D. Umoh, O.J. Peter, F.A. Edogbanya, O. Oguntolu, K. Oshinubi, S. Amadiegwu, Forecasting of COVID-19 pandemic in Nigeria using real statistical data, Commun. Math. Biol. Neurosci. 2021 (2021), Article ID 2.