Statistical analysis of the impact of environmental temperature on the exponential growth rate of cases infected by COVID-19

George Livadiotis*

Southwest Research Institute, San Antonio, TX, United States of America

*glivadiotis@swri.edu

Abstract

We perform a statistical analysis for understanding the effect of the environmental temperature on the exponential growth rate of the cases infected by COVID-19 for US and Italian regions. In particular, we analyze the datasets of regional infected cases, derive the growth rates for regions characterized by a readable exponential growth phase in their evolution spread curve and plot them against the environmental temperatures averaged within the same regions, derive the relationship between temperature and growth rate, and evaluate its statistical confidence. The results clearly support the first reported statistically significant relationship of negative correlation between the average environmental temperature and exponential growth rates of the infected cases. The critical temperature, which eliminates the exponential growth, and thus the COVID-19 spread in US regions, is estimated to be $T_C = 86.1 \pm 4.3 ^\circ F$.

1. Introduction

The daily number of new cases infected by COVID-19 is currently exponentially growing for most countries affected by the virus. However, this exponential growth rate varies significantly for different regions over the globe. It is urgent and timely to understand the reasons behind this regional variation of the exponential growth rates. Little information is known about this matter, while there are indications that the environmental temperature may be a factor; for instance, northern and colder US and Italian regions experienced much more incidents than others.

Typically, the evolution curve of the spread of the coronavirus initiates with a pre-exponential phase, which is characterized by a mild logarithmic growth, followed by the outbreak, that is, the phase of the exponential growth. Social-distancing measures against the spread may affect the evolution curve in a way that the exponential growth slows down (decelerated phase) and starts to decline (decline or decay phase [1]), depending though on the effectiveness and applicability of these measures. However, after the decline of the spread at some place, new infected cases may outbreak in other places, marked with insignificant number of cases until
that moment. Then, a newly growth phase may appear. For example, Fig 1 (left) shows the evolution curve of the spread for the infected cases in mainland China; clearly, we observe the whole growth–decay cycle, as well as, a new re-growth phase.

Super-strict measures, such as complete shut down and quarantines, can successfully lead to the deceleration of the exponential growth of infected cases [2]. Unfortunately, they cannot be successfully applied and followed within vast regions, and especially, for a long and indefinite period of time. Inevitably, measures may be loosened during the decay phase – if not earlier, leading to the birth of an equally disastrous re-growth phase.

The exponential growth is the most effective phase for the evolution curve of infected cases; and the most important question regarding this evolution is still open [3]: What can influence the exponential growth rate, and thus, "flatten the curve"? Measures, strict or not, may affect the evolution of new infected cases, by shifting the spread curve from the exponential to the decelerated growth. It should be noted though that measures do not affect the exponential growth rate itself, but only the period of time that this exponential phase applies. Then, what factors do affect the exponential growth rate?

The age distribution in the place where the outbreak occurs is unlikely to be a factor; indeed, the number of new cases is known to be positively correlated with age, however, the exponential growth rate (China: 0.169; US: 0.121; Italy: 0.090 –decreasing rate) appears to be negatively correlated to the age median of these countries (China: 37.4; US: 38.1; Italy: 45.5 –increasing age); hence, the age is likely irrelevant to the rate variations.

In addition, culture in social activities may be a factor; for example, this might be contributing in the observed differences among the exponential rates in the cases of China, Italy, and US (Fig 1). However, what is causing the major variation of exponential rates among different regions of the same culture? It is apparent that culture does not constitute the main factor influencing the exponential rate.

Fig 2 shows the regional variation of infected cases (left) during the exponential growth phase and the average winter temperature (right) in Italy. The possible negative correlation, observed between regional number of infected cases and winter temperature in Italy, is an indication of the influence of temperature on the exponential growth, but it certainly does not constitute a necessary condition. The reason is that the map plots the total number of the infected cases \( N_t \), which is not dependent only on the exponential growth rate \( \lambda \), but also on the initial number of cases \( N_0 \).

It is generally accepted that the initial infected cases in Italy were travelled directly from China; since some destinations are more favorable than others, then, the initial number of cases \( N_0 \), as well as the current number of cases \( N_t \) (which is proportional to \( N_0 \)), should be subject of regional variation. Therefore, there is a non-negligible possibility, the observed regional variation of the number of infected cases \( N_t \) to be caused by the regional distribution of the initial cases \( N_0 \). In such a case, main airport cities would have incredibly high number of infected cases outplaying a possible negative correlation of daily infected cases with regional average temperature \( T \); the latter may be one of the reasons of the high numbers of cases observed in New York City and Rome.

On the other hand, in their letter to the White House, members of a National Academy of Sciences committee said that "There is some evidence to suggest that [coronavirus] may transmit less efficiently in environments with higher ambient temperature and humidity; however, given the lack of host immunity globally, this reduction in transmission efficiency may not lead to a significant reduction in disease spread without the concomitant adoption of major public health interventions" [4].

Nevertheless, it has to be stressed out that there were no statistical analyses focused on the exponential growth rates of the infected cases in regions with different temperatures. For
instance, several authors (e.g., [5, 6]) found insignificant correlations between temperatures and confirmed cases. However, their analysis was performed on the number of the infected cases $N_t$, which is subject to the randomness of the initial cases $N_0$ as explained above, and not
on the exponential growth rate $\lambda$, which is clearly dependent on physical characteristics of the coronavirus, binding protein, and environment.

Analysis of regional cases can show whether the speculated negative correlation between temperature and number of infected cases is true, meaning a negative correlation between temperature and exponential growth rate. If the environmental temperature plays indeed a substantial role on the virus spread, then, this can provide promising results, such as, the estimation of the critical temperature that may eliminate the number of daily new cases in heavily infected regions.

The purpose of this paper is to improve our understanding of the effect of environmental temperature on the spread of COVID-19 and its exponential growth rate. In particular, we calculate the exponential growth rates of infected cases for US and Italian regions, derive the relationship of these rates with the environmental temperature, evaluate its statistical confidence, and determine the critical temperature that eliminates this rate.

2. Theory

2.1 Modeling behind “flattening the curve”

A standard model for describing the evolution of the infected cases by viruses can be constructed as follows

$$\frac{dx}{dt} = E(x) \cdot I(x), \text{ with } E(x) = \lambda \cdot x \cdot I(x) = 1 - x^b,$$

(1)

where $x(t) = N(t)/N_{\text{max}}$, $x_0 = N_0/N_{\text{max}}$; $N(t)$ is the number of total infected cases evolved from the initial $N_0 = N(0)$ cases, $N_{\text{max}}$ is the maximum possible number of infected case; $\lambda$ is the exponential growth rate, and becomes clear for $x(t) \ll 1$ where $I$ is negligible, leading to:

$$x(t) = x_0 \cdot \exp(\lambda t) \text{ or } N(t) = N_0 \cdot \exp(\lambda t), \text{ for } x \ll 1 \text{ where } I(x) \approx 1.$$  

(2A)

Typically, data of infected cases are daily provided and updated. Thus we set the readout of $N(t)$ on a daily basis, such as: $N_i = N(t/d)$. Hence, we may write

$$x_i = x_0 \cdot \exp(\lambda t) \text{ or } N_i = N_0 \cdot \exp(\lambda t),$$

(2B)

with where $x_0 = N_0/N_{\text{max}}$ with $t$ indicating the time on a daily basis, ($t = 1d, 2d, \ldots$).

The function of negative feedback $I$ models the factors that flattens the curve, such as, the measures taken against spreading. While these factors are not affecting the exponential growth rate $\lambda$, they become more effective as the number of cases increases, getting closer to $N_{\text{max}}$; exponent $b$ controls the effectiveness of these factors; strict (loose) measures correspond to smaller (larger) values of $b$.

Fig 3 shows the evolution curve of the number of new ($\Delta N_i = N_{i+1} - N_i$) and total infected cases ($N_i$), as well as, how this curve flattens for stricter measures (i.e., smaller values of $b$).

As observed in Fig 3(A), stricter measures, nicely modeled by decreasing $b$, do not affect the exponential rate $\lambda$ but they successfully flatten the curve. However, the same can be achieved by downgrading the exponential growth rate, as shown in Fig 3(B).

It is apparent, then, how much useful would it be to know the factors that can flatten the curve by decreasing directly the exponential growth rate, instead of applying stricter measures. Applied measures could be loosen and shorter!

Model (1) originates from the logistic map family (e.g., [7], and references therein; [8]); other complicate versions, such as, the Susceptible–Infectious–Recovered models (e.g., [9]) may be expressed by multi-dimensional differential or difference equations (e.g., [10], and references therein; [11]), but still, the curve flattening is governed by the same features. The two composites, the exponential growth $E$ and the negative feedback $I$, are just the main and
necessary conditions for reproducing the growth-decay phases of the spread curve. Their interplay shows how the spread curve can be flattened as a result of stricter measures, independently of the existent exponential rate.

2.2 Main factors influencing the exponential growth rate

What are the main factors that can affect the exponential growth rate $\lambda$ of COVID-19 spread?

The rate $\lambda$ is expected to have positive correlation with the reproduction number $R_0$ (e.g., proportional to its logarithm), and negative correlation with the incubation period $\tau$ (e.g., inverse proportional) [12]. The number $R_0$ is a measure of how contagious a disease is; it provides the average number of people in a susceptible population that a single infected person will spread the disease over the course of their infection [9], and depends on the physical characteristics of coronavirus [13]. The incubation period $\tau$ is the time elapsed between exposure to coronavirus and first symptoms; during this period, an infected individual cannot infect others; other characteristic periods and time intervals are the latent period between exposure and infection, and the generation time, mostly concerned with transmission process [14].

Characteristic values for COVID-19 are $\tau \sim 5–6$ days and $R_0 \sim 2–4$ [15]. The rate expression can be written as $\lambda \times \ln R_0 / \tau$, and involves all the physical characteristics of the mechanisms of infection and the environmental interactions; this can be easily derived, considering difference equations (that is, iterated discrete maps) (e.g., see: [16–18]). Setting the time to be given in discrete $\tau$-steps ($t = 1\tau, 2\tau, \ldots$), then, by definition of $R_0$ (average number of people that a single infected person will spread the disease), we have $N_t = R_0 N_{t-\tau}$, that is,

$$N_t = R_0 N_{t-\tau} = \cdots = R_0^t N_0,$$

thus,

$$N_t = N_0 \cdot \exp \left( \frac{1}{\tau} \ln R_0 \cdot t \right).$$
We note that the number of the infected cases does not vary for times $t$ taken in-between the integer multiples of $\tau$, but this is not expected in mixtures of populations with random characteristics. Indeed, in a mixture of $M$ evolving infected populations with different initial number of cases $N_0^{(m)}$ and starting times $t_0^{(m)}$, $m = 1, 2, \ldots, M$, the total number of infected cases $N$ at a continuous time $t$ is given by

$$N(t) = \sum_{m=1}^{M} N_0^{(m)} \cdot \exp \left[ \frac{1}{\tau} \ln R_0 \cdot \left( t - t_0^{(m)} \right) \right] = N_0 \cdot \exp \left( \frac{1}{\tau} \ln R_0 \cdot t \right),$$

(3C)

with $N_0 \equiv \sum_{m=1}^{M} N_0^{(m)} \cdot \exp \left( -\frac{1}{\tau} \ln R_0 \cdot t_0^{(i)} \right)$,

(3D)

which coincides with (3b), but with time $t$ varying on a daily basis, independently of the larger value of $\tau \sim 5$ days [15]. Therefore, we set the readout of the total number of infected cases $N$ on a daily basis, such as: $N_t \equiv N(t/[d])$; then, Eq (3D) matches Eq (2B),

$$N_t = N_0 \exp \left( \frac{1}{\tau} \ln R_0 \cdot t \right) \equiv N_0 \exp(\lambda \cdot t),$$

(4)

where the exponential rate is given by:

$$\lambda = \frac{(\ln R_0)}{\tau}.$$

(5)

The main factors that can affect the exponential rate $\lambda$ are: (a) culture in social activities, and (b) environmental temperature and/or other thermodynamic parameters. Intense cultural and social activities have reasonably a positive correlation with $R_0$. As previously mentioned, measures against the virus spread do not effectively influence the exponential growth rate; e.g., they do not change the culture in social activities, which are characteristic of the particular population, but they just cease these social activities for some period of time. In terms of modeling, measures appear only in the negative feedback factor $I$ and not in the $E$ factor of model in Eq (1), while the culture, together with the environmental temperature, are the two main parameters affecting $R_0$ directly. Potentially, the environmental temperature $T$ can affect all the parameters influencing exponential rate. We approach this dependence by (i) a linear approximation of the phenomenological relationship between exponential rate and temperature, and (ii) the connection of reproduction number with Arrhenius behavior (with negative activation energy):

(i) The temperature can affect the physical properties of coronavirus, such as, the incubation time $\tau$, as well as, the reproduction number $R_0$ that depends on these physical properties [13]. A linear approximation absorbs the (weak) temperature dependence of any parameters involved in the exponential rate; then, Eq (5) is linearly expanded as:

$$\lambda(T) \approx \frac{1}{\tau} \ln(R_0^{\text{NTP}}) + |\partial \lambda / \partial T|_{\text{NTP}} \cdot (T^{\text{NTP}} - T) + O(T^2),$$

(6)

where we set the intercept to be given in normal conditions of atmospheric temperature and
pressure (NTP) (that is, $T = 20 \, ^\circ C$, $P = 1\text{atm}$). Then, we rewrite the exponential rate as:

$$\lambda(T) = \lambda_0 \cdot (1 - T_C^{-1} \cdot T),$$  

(7A)

$$\lambda_0 \equiv \frac{1}{T} \ln(R_0^{\text{NTP}}) + |\partial \lambda / \partial T|^{\text{NTP}} \cdot T^{\text{NTP}}, \quad p_2 = -\lambda_0 T_C^{-1} \equiv -|\partial \lambda / \partial T|^{\text{NTP}},$$  

(7B)

where $\lambda_0$ and $p_2$ are the intercept and slope of the linear relation (7a).

(ii) Coronavirus uses their major surface spike protein to bind on a receptor—a protein that acts like a doorway into a human cell [19]. The whole process is a slow chemical reaction, where the mechanism behind can lead to rates negatively correlated with temperature, i.e., increasing rate with decreasing temperature. This is consistent to reaction rate expressed by the Arrhenius exponential with negative activation energy $\exp[|E_a|/k_B T]$ [20]. Then, the effective reproduction number $R_0(T)$ is expressed as a product combining the reproduction number in the absence of temperature effect, $R_0$, and the Arrhenius exponential rate, namely,

$$R_0(T) = R_0^\infty \cdot \exp[|E_a|/(k_B T)], \quad \text{with} \quad R_0^\infty = R_0^{\text{NTP}} \cdot \exp[-|E_a|/(k_B T^{\text{NTP}})].$$  

(8)

Then, Eq (5) gives

$$\lambda(T) = -\frac{1}{T} \frac{|E_a|}{(k_B T^{\text{NTP}})} - \ln(R_0^{\text{NTP}}) + \frac{1}{T} (|E_a|/(k_B)) \cdot T^{-1}.$$  

(9)

We rewrite this expression as:

$$\lambda(T) = \lambda_0 \cdot (-1 + T_C \cdot T^{-1}),$$  

(10A)

$$\lambda_0 \equiv \frac{1}{T} \frac{|E_a|}{(k_B T^{\text{NTP}})} - \ln(R_0^{\text{NTP}}), \quad p_2 = \lambda_0 T_C^{-1} \equiv |E_a|/(T k_B),$$  

(10B)

where $-\lambda_0$ and $p_2$ are the intercept and slope of the linear relation (10a), respectively.

Reactions of negative activation energy are barrier-less, relying on the capture of the molecules in a potential well. Increasing [decreasing] the temperature leads to a reduced [gained] probability of the colliding molecules capturing one another. Due to the negative activation energy, decreasing the environmental temperature reduces the probability of virus-protein reaction, thus the virus may stay inactive on air or surfaces and eventually die.

Exponential growth is related to community spread through outdoors activities, while the decelerated growth caused by effective measures is related to indoors activities: Exponential growth exists once the disease is still effective and the measures are loosened, allowing people to outdoor social activities; however, exponential growth decelerates followed by the decay phase, once effective measures hold people in small groups indoors. Therefore, exponential growth rate must be related to outdoors (rather than indoors) activities, and thus to the environmental temperature. As long as the exponential growth takes place, the environmental temperature has an effective role on the chemical reaction between virus and spike protein.

It should be noted that both the models (7b) and (10b) consider that the exponential rate $\lambda$, or the reproduction number $R_0$, are subject to a component influenced by the culture in social activities (intercept $\lambda_0$) and a component mostly influenced by the temperature (linear term with slope $p_2$). In this way, the slope may indicate some universal quantity involved, such as, the (negative) activation energy.
Next, we employ the above two expressions of exponential rate $\lambda$ and temperature $T$, Eqs (7A and 10A), in order to set the two types of statistical models for fitting $(T, \lambda)$ measurements for US and Italian regions.

### 3. Methods

#### 3.1 Data

We use publicly available datasets of: (1) average environmental temperature of US and Italian regions (e.g., see: www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/climate-normals; it.climate-data.org; www.weather-atlas.com); (2) time series of the number of daily infected cases of US and Italian regions (e.g., see: www.thelancet.com; www.protezionecivile.gov.it).

#### 3.2 Data analysis

We analyze the datasets of regional infected cases in US and Italy, derive the relationship of the exponential growth rate of the number of cases with temperature, and evaluate its statistical confidence. First, we derive the exponential growth rates of the infected cases characterizing each examined region of US and Italy; then, we plot these values against the environmental temperatures of each region, and perform the corresponding statistical analysis. We proceed according to the following steps:

i. Collect the time series of the current infected cases $N_t$ for all US and Italian regions.

ii. For each of the US and Italian regions, we plot log ($N_t$) and log ($\Delta N_t$) with time $t$, detect the time intervals of linear relationship corresponding to the phase of exponential growth, fit the data-points within this region, and derive the slope (on linear-log scale), that is, the exponential growth rate $\lambda$. The total $N_t$ and new cases $\Delta N_t$ should be characterized by the same exponential rate, $\lambda$, thus the slopes resulted from the linear fits of log ($N_t$) and log ($\Delta N_t$) with time are (weighted) averaged (Fig 4).

iii. Collect environmental temperature data, and calculate the temperature averaged over the whole examined region. The incubation period $\tau$ is longer than the time scale of a single day or night, thus the temperature is averaged over the daily and nightly measurements.

iv. Co-plot all the derived sample values $(T \pm \delta T, \lambda \pm \delta \lambda)$, where each pair corresponds to each examined region; then, apply a linear fitting in order to derive the linear relationship between $T$ and $\lambda$, as well as, evaluate the statistical confidence of this relationship; repeat the same for all US and Italian regions.

v. Determine the critical temperature $T_C$ for which the rate becomes negligible; to eliminate the uncertainties of $T_C$ as a fitting parameter, we perform the linear fitting with the statistical model $\lambda = \lambda_0 (1 - T/T_C)$ instead of $\lambda = p_1 + p_2 T$.

vi. Repeat (iv) and (v) with pairs of $(T^{-1} \pm \delta T^{-1}, \lambda \pm \delta \lambda)$; we estimate again $T_C$ by performing the linear fitting with the statistical model $\lambda = \lambda_0 (-1 + T_C/T)$ instead of $\lambda = p_1 + p_2 T^{-1}$.

#### 3.3 Statistical analysis

The hypothesis to be tested is that the exponential growth rate $\lambda$ varies linearly with temperature; ($x$ is set to be the temperature or its inverse). This is tested by examining the chi-square corresponding to the fitting of the two-parameter linear statistical model $\lambda(x; p_1, p_2) = p_1 + p_2 x$ to
Fig 4. Linear fitting of the number of the total \( N_t \) and new \( \Delta N_t \) infected cases with time (on linear-log scale) for the states of California and Illinois, where the slope reads the exponential rate \( \lambda \). The resulted rates from the linear fitting of \( \log (N_t) \) (black) and \( \log (\Delta N_t) \) (red) are (weighted) averaged (blue). The phases are color-coded as in Fig 1.

The resulted rates from the linear fitting of \( \log (N_t) \) (black) and \( \log (\Delta N_t) \) (red) are (weighted) averaged (blue). The phases are color-coded as in Fig 1.

https://doi.org/10.1371/journal.pone.0233875.g004

the given \( N \) data points, (The number of data point, \( N \), should not to be confused with the number of cases, \( N_t \)). Therefore, we minimize the chi-square \( \chi^2(p_1, p_2) = \sum_{i=1}^{N} \sigma_i(p_2)^{-2} \)

\[ \lambda_i - p_1 - p_2 x_i \right)^2, \]

where the total variance that characterize each data point is now given by \( \sigma_i(p_2)^{-2} = \sigma_{i,1}^2 + p_2^2 \sigma_{i,2}^2 \) [21]. The global minimum of the chi-square function \( \chi^2(p_1, p_2) \) gives the optimal parameter values, \( (p_1^*, p_2^*) \), by solving the normal equations \( \partial \chi^2(p_1, p_2)/\partial p_1 = 0 \) and \( \partial \chi^2(p_1, p_2)/\partial p_2 = 0 \); the minimum chi-square value is \( \chi^2_{\text{min}} = \chi^2(p_1^*, p_2^*) \). The statistical errors of these values are given by \( \delta_{p_1} = \sqrt{H_{11}^{-1}}, \delta_{p_2} = \sqrt{H_{22}^{-1}}, \alpha = 1,2 \), where \( H \) is the Hessian matrix of the chi-square at the global minimum, and \( H_{\alpha \alpha} \) is the \( \alpha \)-th diagonal element of its inverse matrix [22,23]; \( \chi^2_{\text{min}} = \chi^2_{\text{red}} = M \chi^2_{\text{red}} \) is the reduced chi-square value for degrees of freedom (dof) equal to \( M = N-2 \). The propagation errors of the measurements \( \{ (x_i \pm \sigma_{i,1}, \lambda_i \pm \sigma_{i,2}) \}_1^N \) are given by

\[ \delta_{p_1} = \sqrt{\left( \frac{\partial \chi^2}{\partial x_i} \right)^2 \sigma_{i,1}^2 + \left( \frac{\partial \chi^2}{\partial \lambda_i} \right)^2 \sigma_{i,2}^2}, \alpha = 1,2, \]

where the derivatives are numerically derived.

We will use two linear statistical models, (a) \( \lambda(T) = \lambda_0(1 - T_a^{-1} - T) \), and (b) \( \lambda(T^{-1}) = \lambda_0(-1 + T_a^{-1} T) \), as given by (Eqs 7A and 10A); both can be written with the linear expression:

\[ \lambda(x; p_1, p_2) = p_1 + p_2 \cdot x, \]  \( 11A \)

(i) \( x = T \), \( p_1 = \lambda_0 \), \( p_2 = -\lambda_0 T_a^{-1} \), and (ii) \( x = T^{-1} \), \( p_1 = -\lambda_0 \), \( p_2 = \lambda_0 T_a \). \( 11B \)

The statistical confidence of the dependence of the exponential growth rate on the environmental average temperature may be sufficiently high, leading to the acceptance of any of the two statistical models. The goodness of the fitting of each model is evaluated using two types of statistical tests, the "reduced chi-square", the "\( p \)-value of the extremes", and their combination (e.g., [24–26]), while Student’s t-test is also used for evaluating the statistical confidence of the derived slopes:
• **Reduced Chi-Square:** The goodness of fitting is estimated by the reduced chi-square value, $\chi^2_{\text{red}} = \frac{1}{M} \chi^2_{\text{min}}$. The meaning of $\chi^2_{\text{red}}$ is the portion of $\chi^2_{\text{min}}$ that corresponds to each of the dof, and $\chi^2_{\text{red}}$ has to be ~1 for a good fit. Therefore, fitting is characterized as "good" when $\chi^2_{\text{red}} < 1$, otherwise there is an overestimation, $\chi^2_{\text{red}} > 1$, of the errors. One order of magnitude less, $\chi^2_{\text{red}} = 0.1$, or more, $\chi^2_{\text{red}} = 10$, can be set as the accepted limits, i.e., $0.1 \leq \chi^2_{\text{red}} \leq 10$.

• **P-value of the extremes:** The goodness of fitting is evaluated by comparing the estimated minimized chi-square value, $\chi^2_{\text{min}}$, and the chi-square distribution,

$$P(\chi^2; M) = \frac{1}{\Gamma(M/2)} (\chi^2)^{M/2-1} e^{-\chi^2/2}$$

that is, the distribution of all the possible $\chi^2$ values (parameterized by the dof = M). The likelihood of having a $\chi^2$ value, equal to or larger than the estimated value $\chi^2_{\text{min}}$, is given by the complementary cumulative distribution. The probability of taking a result $\chi^2$, larger than the estimated value $\chi^2_{\text{min}}$, defines the $p$-value that equals

$$P(\chi^2_{\text{min}} \leq \chi^2 < \infty) = \int_{\chi^2_{\text{min}}}^{\infty} P(\chi^2; M) d\chi^2$$

The larger the $p$-value, the better the fitting. According to this method, the probability of taking a result with $\chi^2$ being extremer than the observed value $\chi^2_{\text{min}}$, defines the $p$-value of the extremes; this equals the minimum between the two probabilities, $P(0 \leq \chi^2 \leq \chi^2_{\text{min}})$, and its complementary, $P(\chi^2_{\text{min}} < \chi^2 < \infty)$. Fits associated with $p$-values smaller than the significance level of 0.05 are typically rejected.

• **Combined P-value and Chi-Square:** The $p$-value of the extremes has very similar behavior with the reduced chi-square [27, 28], because, (i) the $p$-value attains the optimal value ($p = 0.5$) when chi-square does ($\chi^2_{\text{red}} = 1$), (ii) larger values of $1 - 2p$ corresponds to larger values of $|\chi^2_{\text{red}} - 1|$, (iii) both fractions $(1-2p)/(1+2p)$ and $|1 - \chi^2_{\text{red}}/(1 + \chi^2_{\text{red}})$ range from 0 to 1, reduced to 9/11 when the accepted limits, $p = 0.05$ or 0.1 $\leq \chi^2_{\text{red}} \leq 10$, are reached. Then, a combined measure can be defined by the sum of the squares of these fractions, i.e.,

$$\sqrt{[(1 - 2p)/(1 + 2p)]^2 + [(1 - \chi^2_{\text{red}})/(1 + \chi^2_{\text{red}})]^2}.$$

• **Student’s t-test:** This is another test for evaluating the statistical significance of the slope derived from the linear fitting of the temperature-rate sample points $(T_i, \pm \delta T_i; \lambda, \pm \delta \lambda_i)$. We examine, whether the slope $p_2 \pm \delta p_2$ has significant difference from the zero slope (null hypothesis: slope is zero), by performing the Student’s t-test with $t_m = p_2/\delta p_2$, where the corresponding $p$-value is derived from the integration of t-distribution

$$P(t; M) = \frac{\Gamma[(M+1)/2]}{\sqrt{\pi M/2}} [(1 + t^2/M)^{-\frac{1}{2}(M+1)}]$$

for $t \leq t_m < \infty$, i.e., $P(t_m; M) = \int_{t_m}^{\infty} P(t; M) dt$. The Student’s t-test is not passed for the null hypothesis that the examined slope equals zero, when the corresponding $p_t$-value is smaller than the acceptable confidence limit of 0.05; then, the null hypothesis is rejected, imposing that the slope has statistically significant difference for zero. In addition, we compare the slopes estimated for US with those estimated for Italian regions, by deriving $t_m = |b_{US} - b_{IT}|/\sqrt{\sigma_{US}^2 + \sigma_{IT}^2}$ and $M_m = (\sigma_{US}^2 + \sigma_{IT}^2)^{\frac{1}{2}} / (\sigma_{US}^2/M_{US} + \sigma_{IT}^2/M_{IT})$, and then, finding again $p_t(t_m; M_m)$; the t-test is passed for the null hypothesis that the examined slopes are equal, when the corresponding $p_t$-value is larger than 0.05; in this case, the null hypothesis is accepted, thus there is no statistically significant difference between the two slopes.

4. Results

The linear fitting of log ($N_t$) or log ($\Delta N_t$) with respect to time $t$ within the region of exponential growth phase, resulted to the respective rates (which are given by the fitted slopes); their weighted averages are shown in Table 1 for US and in Table 2 for Italian regions, while plotted
against the average regional temperature in Figs 5 and 6, respectively. The method of weighted fitting for double uncertainties \((x_i \pm \delta x_i, \lambda_i \pm \delta \lambda_i)\), as described by [21], is used for estimating the fitting parameters \(\lambda_0, T_C\), together with their statistical, propagation, and total errors. The fits of the linear statistical model with temperature, \(x_i = T_i\) (left panels in Figs 5 and 6), as well as of the alternative statistical model with inverse temperature, \(x_i = T_i^{-1}, \delta x_i = \delta T_i T_i^{-2}\), (right panels in Figs 5 and 6), are both characterized with high statistical confidence, attaining high \(p\) -values (>0.05) and reduced chi-squares \(\chi^2_{\text{red}}\) values (close to 1); also, both fits provide similar estimations of \(T_C\). The fitting results are shown in Table 3.

We also examine whether the sample points \((T_i \pm \delta T_i, \lambda_i \pm \delta \lambda_i)\) are subject to statistically significant concentrations or rarefactions, namely, whether possible heterogeneities within the distribution of sample points plays significant role in the fitted relationship. For this, we derive the temperature-rate relationship and its statistical confidence by fitting the homogenized set of sample points, instead of the raw sample points; then, we examine whether the fitting parameters differ from those derived from fitting the raw sample points. We homogenize the sample points by grouping them in temperature bins of \(\Delta T \sim 1\) C (e.g., see: [29]). We estimate the weighted mean and error of the rates included in each bin. In the case of US regions we

| Region | \(T_0\) [C] | \(\delta T\) [C] | \(\lambda\) [d\(^{-1}\)] | \(\delta \lambda\) [d\(^{-1}\)] |
|--------|-----------|-------------|-------------|-------------|
| MI     | -1.23     | 4.2         | 0.1709      | 0.0096      |
| WI     | -0.7      | 2.6         | 0.1704      | 0.0099      |
| MA     | 1.7       | 3.5         | 0.1495      | 0.0183      |
| CT     | 2.9       | 4.0         | 0.1463      | 0.0062      |
| PA     | 3.1       | 2.9         | 0.1349      | 0.0019      |
| WA     | 3.7       | 3.5         | 0.1432      | 0.0162      |
| NJ     | 4.5       | 3.9         | 0.1523      | 0.0114      |
| IN     | 5.1       | 5.2         | 0.1237      | 0.0064      |
| OH     | 5.1       | 4.2         | 0.1256      | 0.0092      |
| IL     | 5.2       | 4.0         | 0.1482      | 0.0050      |
| CO     | 5.6       | 4.6         | 0.1108      | 0.0200      |
| NY     | 5.8       | 2.8         | 0.1203      | 0.0096      |
| MO     | 6.5       | 2.8         | 0.1322      | 0.0079      |
| VA     | 7.4       | 2.8         | 0.0946      | 0.0015      |
| TN     | 8.8       | 5.2         | 0.1266      | 0.0098      |
| NC     | 9.7       | 1.9         | 0.1398      | 0.0073      |
| SC     | 11.3      | 2.8         | 0.1027      | 0.0044      |
| GA     | 13.3      | 2.8         | 0.1169      | 0.0096      |
| CA     | 14.2      | 3.9         | 0.0908      | 0.0017      |
| LA     | 15.2      | 4.2         | 0.1081      | 0.0122      |
| TX     | 15.3      | 5.7         | 0.1083      | 0.0070      |
| AZ     | 17.0      | 2.7         | 0.0786      | 0.0087      |
| FL     | 19.5      | 4.6         | 0.1033      | 0.0072      |

(1) The given values of exponential growth rate and its uncertainty is the result of the weighted averaging of the rates derived from total and new infected cases; (2) the environmental temperature is averaged over the time period, from \(\tau \sim 5\) days before the appearance of the 1\(^{st}\) case, to 1\(^{st}\) April; (3) the standard deviation of temperature is given by the half difference between highest and lowest values within the examined time period, divided by \(\sqrt{2}\) (similar to the standard deviation for a sinusoidal function); (4) NY: The temperature refers to the New York City, instead of the whole state, which suffers from the vast majority of the state infected cases.

https://doi.org/10.1371/journal.pone.0233875.t001
also performed a homogenization of rates, by grouping the temperature-binned means in rate bins of $\Delta \lambda \sim 0.01 \text{ d}^{-1}$. In the case of sample points with inverse temperatures, $(T_i^{-1} \pm \delta T_i^{-1}, \lambda_i \pm \delta \lambda_i)$, the procedure is exactly the same. Homogenized datasets result in a smooth relationship between the values of binned temperature and rate, as it can be observed in the plots of rate against temperature or inverse temperature (left or right lower panels, respectively), and for both US and Italy regions (Figs 5 and 6, respectively). The results are highly supportive of the negative correlation between rate and temperature. The results are shown in Table 4. We observe that the linear relationships of the growth rate with temperature or inverse temperature are characterized by high statistical confidence for the homogenized datasets ($p$-values much higher than the significant limit of 0.05; $\chi^2_{\text{red}}$ far from the significant limits of 0.1 and 10). Therefore, the arrangement of sample points do not affect significantly the fitting results.

In addition, as shown in Tables 3 and 4, the linear fits of sample points $(T_i \pm \delta T_i, \lambda_i \pm \delta \lambda_i)$ and $(T_i^{-1} \pm \delta T_i^{-1}, \lambda_i \pm \delta \lambda_i)$ do not pass the Student’s $t$-test for the null hypothesis that their slopes equals zero, i.e., the corresponding $p_r$-values are smaller than the acceptable confidence limit of 0.05; therefore, the negative correlation of environmental temperature with the exponential rate is statistically significant (accepted with confidence 95%).

In order to improve the statistics of the estimated critical temperature, we combine the sample points $(T_i \pm \delta T_i, \lambda_i \pm \delta \lambda_i)$ of US and Italian regions. First, we perform the Student’s $t$-test to compare the slopes from these regions; we find high $p_r$-values ($>0.05$) for both fits of $x = T$ and $x = T^{-1}$, thus, the two populations are likely characterized by the same slope. The respective intercept $\lambda_0$ does not pass the same test, i.e., the intercepts corresponding to US and Italian regions are likely different; (that is expected, given of the different culture). A universality may characterize the slopes of two countries, either for the fits with $x = T$ or $x = T^{-1}$, i.e., $p_2 = (\partial \lambda / \partial T)^{\text{NTP}}$ or $p_2 = |E_a|/(rK_B)$, respectively.

### Table 2. Averaged temperatures and estimated exponential rates of Italian regions.

| Region               | $T$  | $\delta T$ | $\lambda$ | $\delta \lambda$ |
|----------------------|------|------------|------------|------------------|
| Aosta Valley         | -4.0 | 2.8        | 0.1187     | 0.0041           |
| S Tyrol              | 2.2  | 2.7        | 0.0887     | 0.0080           |
| Abruzzo              | 4.8  | 3.1        | 0.1055     | 0.0127           |
| Piedmont             | 6.2  | 3.1        | 0.0983     | 0.0068           |
| Molise               | 7.5  | 3.2        | 0.0667     | 0.0158           |
| Basilicata           | 7.5  | 3.2        | 0.0621     | 0.0141           |
| Friuli Venezia Giulia| 7.5  | 1.8        | 0.0816     | 0.0147           |
| Veneto               | 7.6  | 2.5        | 0.0674     | 0.0027           |
| Liguria              | 7.7  | 2.4        | 0.0693     | 0.0106           |
| Tuscany              | 7.8  | 3.0        | 0.0900     | 0.0127           |
| Umbria               | 8.1  | 3.1        | 0.0900     | 0.0127           |
| Lazio                | 11.1 | 2.4        | 0.0906     | 0.0065           |
| Lombardy             | 8.1  | 2.9        | 0.0942     | 0.0089           |
| Emilia Romagna       | 8.3  | 3.0        | 0.0695     | 0.0095           |
| Marche               | 10.4 | 2.0        | 0.0696     | 0.0140           |
| Calabria             | 10.5 | 2.5        | 0.0939     | 0.0155           |
| Sicily               | 11.0 | 2.3        | 0.0762     | 0.0053           |
| Campania             | 11.6 | 2.0        | 0.0604     | 0.0076           |

Similar to the notes (1)-(3) of Table 1.

https://doi.org/10.1371/journal.pone.0233875.t002
Next, we perform the linear fits of the sample points \((T_i \pm \delta T_i, \lambda_i \pm \delta \lambda_i)\) for the mixed set of US and Italian data, once the rates of the Italian regions are shifted by \(\Delta \lambda\); (this is allowed, since it has just be shown that a universality is likely characterized the slopes). The optimal fitting is obtained for that shift \(\Delta \lambda\), for which the reduced chi-square is \(\sim 1\), the \(p\)-value of the extremes is \(\sim 0.5\), and the combined measure \(\sim 0\) (see previous section).

**Fig 5.** Linear fitting of rates with (left) temperatures and (right) inverse temperatures for US regions. The fitting is weighted with double uncertainties (i.e., on both the temperature and rate values). The analysis is first completed for the raw measurements (upper) and then repeated for the binned averages (lower).

https://doi.org/10.1371/journal.pone.0233875.g005

Table 5 includes the weighted means of slopes for the fits \(x = T\) or \(x = T^{-1}\), with slopes \(p_2 = -[\partial \lambda_i/\partial T]^{NTP}\) and \(p_2 = |E_a|/(\tau k_B)\), respectively. The latter can be used for deriving the activation
energy, where \( \tau = 5.2 \pm 1.1 \text{ d} [15] \), where we find \( |E_a| = 0.212 \pm 0.057 \text{ eV} \). The value of the critical temperature is \( T_C = 303.2 \pm 2.4 \text{ K} \) (86.1 ± 4.3 F° or 30.1 ± 2.4 C°). In addition, using (Eqs 7B and 7D).

Fig 6. As in Fig 5, but for Italian regions.

https://doi.org/10.1371/journal.pone.0233875.g006

Table 3. Fitting parameters of temperature–rate values for US and Italian regions.

| US/Model | \( \lambda_0 \) [d⁻¹] | \( \delta \lambda_0 \) [d⁻¹] | \( p_2 \) * | \( \delta p_2 \) | \( T_C \) [C°] | \( \delta T_C \) [C°] | \( \chi^2_{\text{red}} \) | \( p \)-value | \( p_t \)-value |
|----------|-----------------|-----------------|-------------|-------------|---------------|---------------|----------------|-------------|-----------------|
| \( x = T, p_1 = \lambda_0, p_2 = -\lambda_0 T_C^{-1} \) | 0.1688 | 0.0125 | -0.00554 | 0.00133 | 30.4 | 5.4 | 0.85 | 0.34 | 2.2×10⁻⁴ |
| \( x = T, p_1 = -\lambda_0, p_2 = \lambda_0 T_C \) | 1.455 | 0.288 | 444 | 107 | 32.0 | 4.8 | 0.84 | 0.32 | 2.3×10⁻⁴ |
| Italy/Model | \( \lambda_0 \) [d⁻¹] | \( \delta \lambda_0 \) [d⁻¹] | \( p_2 \) | \( \delta p_2 \) | \( T_C \) [C°] | \( \delta T_C \) [C°] | \( \chi^2_{\text{red}} \) | \( p \)-value | \( p_t \)-value |
| \( x = T, p_1 = -\lambda_0, p_2 = \lambda_0 T_C \) | 0.1173 | 0.0169 | -0.00465 | 0.00206 | 25.3 | 7.8 | 0.95 | 0.49 | 0.019 |
| \( x = T, p_1 = -\lambda_0, p_2 = -\lambda_0 T_C^{-1} \) | 1.20 | 0.47 | 361 | 166 | 25.3 | 7.8 | 0.95 | 0.49 | 0.019 |

* Units of the slope \( p_2 \) are [d⁻¹K⁻¹] when \( x = T \), and [d⁻¹K] when \( x = T^{-1} \).

https://doi.org/10.1371/journal.pone.0233875.t003
we derive the reproduction number \( R_0 \), i.e.,

\[
\frac{1}{\tau} \ln(R_{0}^{\text{NTP}}) = \lambda_0^{x=T} - \lambda_0^{x=T} \cdot (T_{\text{NTP}} / [C^0]) - 1, \quad \text{or} (12A)
\]

The two formulae in Eq (12B) provide the value of \( \frac{1}{\tau} \ln(R_{0}^{\text{NTP}}) \) as 0.0572 ± 0.0098 and 0.0534 ± 0.0146, respectively, with weighted mean 0.0560 ± 0.0084; then, we find \( R_{0}^{\text{NTP}} \approx 1.34 \pm 0.10 \), that is, the reproduction number for \( T = 20^\circ C \). The corresponding number at \( T = 0^\circ C \) is \( R_0(0^\circ C) \approx 2.47 \pm 0.45 \), while by substituting the estimated parameters in Eq (8), we derive the

\[
\chi^2_{\text{red}} \quad p\text{-value} \quad p_1 \text{-value}
\]

| Model | \( \lambda_0 \) [d\(^{-1}\)] | \( \Delta \lambda_0 \) [d\(^{-1}\)] | \( p_2 \) | \( \Delta p_2 \) | \( T_C [C^0] \) | \( \Delta T_C [C^0] \) | \( \chi^2_{\text{red}} \) | p-value | p\(_1\)-value |
|-------|------------------|------------------|----------|-----------|------------------|------------------|------------------|------------|------------|
| \( x = T, p_1 = \lambda_0, p_2 = -\lambda_0 T_C^{-1} \) | 0.1645 | 0.0089 | -0.00488 | 0.00096 | 33.7 | 5.1 | 0.16 | 1.9x10\(^{-3}\) |
| \( x = T^1, p_1 = -\lambda_0, p_2 = \lambda_0 T_C \) | 1.331 | 0.2417 | 409 | 81 | 34.4 | 4.9 | 0.16 | 0.021 |
| \( \text{Italy/ Model} \) | \( \lambda_0 \) [d\(^{-1}\)] | \( \Delta \lambda_0 \) [d\(^{-1}\)] | \( p_2 \) | \( \Delta p_2 \) | \( T_C [C^0] \) | \( \Delta T_C [C^0] \) | \( \chi^2_{\text{red}} \) | p-value | p\(_1\)-value |
| \( x = T, p_1 = \lambda_0, p_2 = -\lambda_0 T_C^{-1} \) | 0.1086 | 0.0127 | -0.00420 | 0.00154 | 25.9 | 6.8 | 1.34 | 0.24 | 2.0x10\(^{-3}\) |
| \( x = T^1, p_1 = -\lambda_0, p_2 = \lambda_0 T_C \) | 1.0374 | 0.2889 | 313 | 115 | 28.3 | 5.8 | 1.28 | 0.27 | 0.021 |

* Same units as in Table 3.

Table 4. Fitting parameters of binned temperature–rate values for US and Italian regions.

Fig 7. Fitting of datasets combined for US and Italian regions, with the latter’s rates shifted by \( \Delta \lambda \). The optimal fitting corresponds to shifts \( \Delta \lambda = 0.031 \) and \( -0.058 \), for which the reduced chi-square is \( -1 \), the p-value of the extremes is \( -0.5 \), and the combined measure \( -0 \).
Table 5. Fitting parameters of combined US and optimally shifted Italian regions.

| $\Delta \lambda$ [d$^{-1}$], x = $T$ | $\lambda_0$ [d$^{-1}$] | $\delta \lambda_0$ [d$^{-1}$] | $p_2$ [d$^{-1}$K$^{-1}$] | $\delta p_2$ [d$^{-1}$K$^{-1}$] | $T_C$ [C$^0$] | $\delta T_C$ [C$^0$] | $T_C$ [F$^0$] | $\delta T_C$ [F$^0$] |
|--------------------------------------|-----------------------|----------------------------|------------------------|------------------------|----------------|----------------|----------------|----------------|
| 0.03098                             | 0.1631                | 0.0116                     | 0.005717               | 0.00136               | 28.538         | 4.97058        | 83.3684        | 8.947044       |
| 0.05766                             | 0.1781                | 0.0113                     | 0.005853               | 0.00126               | 30.429         | 4.85595        | 86.7722        | 8.74071        |
| Weighted Mean                       | 0.1708                | 0.0110                     | 0.00579                | 0.00093               | -              | -              | -              | -              |
| $\Delta \lambda$ [d$^{-1}$], x = $T^1$ | $\lambda_0$ [d$^{-1}$] | $\delta \lambda_0$ [d$^{-1}$] | $p_2$ [d$^{-1}$K$^{-1}$] | $\delta p_2$ [d$^{-1}$K$^{-1}$] | $T_C$ [C$^0$] | $\delta T_C$ [C$^0$] | $T_C$ [F$^0$] | $\delta T_C$ [F$^0$] |
| 0.02926                             | 1.531                 | 0.303                      | 463                    | 113                   | 29.4           | 4.3            | 84.9           | 7.8            |
| 0.05899                             | 1.575                 | 0.282                      | 480                    | 107                   | 31.5           | 4.2            | 88.8           | 7.6            |
| Weighted Mean                       | 1.555                 | 0.208                      | 472                    | 78                    | 30.1$^*$       | 2.4            | 86.1           | 4.3            |

* The weighted mean of $T_C$ in [C$^0$] or [F$^0$] takes into account all four estimated values.

https://doi.org/10.1371/journal.pone.0233875.t005

Fig 8. Relationship of the reproduction number $R_0$ and its uncertainty with environmental temperature $T$. According to this, new affected cases cease ($R_0 = 1$) when temperature climbs to $T_C \approx 30$ C$^0$ or (~86 F$^0$).

https://doi.org/10.1371/journal.pone.0233875.g008
A general relationship for any temperature (expressed in K), also plotted in Fig 8:

$$ R_0(T|C^0) = (1.334 \pm 0.10) \cdot \exp\left(2450 \pm 660(T^{-1} - 293.15^{-1})\right). \tag{13} $$

5. Discussion and conclusions

Up-to-date there is no systematic statistical analysis of the effect of the environmental temperature $T$ (and possibly other weather parameters) on the exponential growth rate of the cases infected by COVID-19, while a statistically confident relationship between temperature and growth rate (either with positive or negative correlation) was unknown.

The presented analysis led to the first statistically confident relationship of negative correlation between the exponential growth rate and the average environmental temperature, derived for US and Italian regions. In particular, we analyzed datasets of regional infected cases in US and Italy, derived the exponential growth rates for each of these regions and plotted them against environmental temperatures averaged within the same regions, derived the relationship of temperature—growth rate, and evaluated its statistical confidence.

The performed statistical analysis involved fitting of linear statistical models with the datasets of environmental temperature (or its inverse) and exponential growth rate. The two linear models developed and used for the statistical analysis are (a) $\lambda(T) = \lambda_0(1-T_C^{-1}T)$, and (b) $\lambda(T^{-1}) = \lambda_0(-1+T_C^{-1}T^{-1})$. The statistical confidence of fitting was evaluated using the reduced chi-square values, the $p$-value of extremes, and a testing measure that combines both of these values; also, the Student’s $t$-test was used to compare the derived slopes.

The sample points of temperature (or inverse temperature) and exponential growth rate were also tested for statistically significant concentrations or rarefactions, that is, for possible heterogeneities within the distribution of sample points that could have significant role in the results. The statistical analysis of the homogenized temperature-rate data points concluded that the negative correlation between temperature and exponential rate is stable, having no statistically significant variability due to concentrations or rarefactions, and it is characterized by a high statistical confidence.
We also performed a Student’s t-test and ensured that the difference between the sample means of US and Italian regions is not statistically significant. A universality is likely characterizing the slope of the temperature-rate relationship. This verifies the modeling developed and used by this analysis, where the exponential rate $\lambda$, or the reproduction number $R_0$, are subject to a component influenced by the culture in social activities (intercept $\lambda_0$) and a component influenced by the temperature (slope $p_2$). In this way, the slope may indicate to a universal quantity involved, such as, the (negative) activation energy.

Having shown that the derived slopes for US and Italian regions are characterized by no statistically confident difference, we improved the statistics of the estimated fitting parameters by combined the sample points of US and Italian regions. From the derived relationship, among others, we were able to estimate the values of the (negative) activation energy $E_a$ as well as the reproduction number $R_0$ at normal conditions and how this depends on temperature.

Therefore, the results clearly showed that there is indeed statistically significant negative correlation of temperature on the exponential growth rate of the cases infected by COVID-19. Fig 9 shows the anti-correlation between the mapped exponential rates and average environmental temperature of the US regions examined by this analysis, which they are characterized by a readable exponential growth phase in their evolution spread curve.
Given the negative correlation of the environmental temperature with the exponential growth rate, it was reasonable to ask for the critical temperature that eliminates the exponential rate, and thus the number of daily new cases in infected regions. This was found to be $T_C \approx 86.1 \pm 4.3$ F for US regions. It is straightforward to ask when the environmental temperature will climb above this critical value. As an example, Fig 10 plots the daily average temperatures in San Antonio, Texas, shown that it will be clearly above the estimated $T_C$ threshold by the end of May.

The resulted high statistical confidence of the negative correlation of the environmental temperature on the exponential growth rate of the cases infected by COVID-19 is certainly encouraging for loosening super-strict social-distancing measures, at least, during the summery high temperatures. However, we are, by no-means, recommending a return-to-work date based only on this study. But we do think that this should be part of the decision, as well as an inspiration for repeating the same analysis in other heavily infected regions. The steps of these analyses may be followed as:

i. Identify different outbreaks in regions with the same culture in social activities and different environmental temperature;

ii. Estimate the exponential growth rates for these regions from the time series of infected cases;

iii. Plot the derived rates against the environmental temperature averaged for these regions, and repeat the analysis of this study to determine the temperature-rate relationship and its statistical confidence.

**Author Contributions**

**Conceptualization:** George Livadiotis.

**Data curation:** George Livadiotis.

**Formal analysis:** George Livadiotis.

**Funding acquisition:** George Livadiotis.

**Investigation:** George Livadiotis.

**Methodology:** George Livadiotis.

**Project administration:** George Livadiotis.

**Resources:** George Livadiotis.

**Software:** George Livadiotis.

**Supervision:** George Livadiotis.

**Validation:** George Livadiotis.

**Visualization:** George Livadiotis.

**Writing – original draft:** George Livadiotis.

**Writing – review & editing:** George Livadiotis.

**References**

1. Smith A. M., Adler F. R., and Perelson A. S. An accurate two-phase approximate solution to an acute viral infection model. J Math Biol. 2010; 60: 711–726.
2. Giugliano F. Greece shows how to handle the crisis. Bloomberg Opinion. 2020; https://www.bloomberg.com/opinion/articles/2020-04-10/greece-handled-coronavirus-crisis-better-than-italy-and-spain.

3. Black A., Liu D., and Mitchell L. How to flatten the curve of coronavirus, a mathematician explains. Medical Express. 2020; https://theconversation.com/how-to-flatten-the-curve-of-coronavirus-a-mathematician-explains-133514.

4. Relman D. A. Rapid Expert Consultation on SARS-CoV-2 Survival in Relation to Temperature and Humidity and Potential for Seasonality for the COVID-19 Pandemic. The National Academies Press. 2020; https://doi.org/10.17226/25771

5. Pawar S., Stanam A., Chaudhari M., and Rayudu D. Effects of temperature on COVID-19 transmission. MedRxiv. 2020; https://doi.org/10.1101/2020.03.29.20044461

6. Yao Y., Pan J., Liu Z., Meng X., Wang W., and Kan H., et al. No Association of COVID-19 transmission with temperature or UV radiation in Chinese cities. Eur. Respir. J. 2020; https://doi.org/10.1183/13993003.00517–2020

7. Livadiotis G. Numerical approximation of the percentage of order for one-dimensional maps. Adv. Complex Sys. 2005; 8:15–32.

8. Wu K., Darcet D., Wang Q., and Sornette D. Generalized logistic growth modeling of the COVID-19 outbreak in 29 provinces in China and in the rest of the world. MedRxiv. 2020; https://doi.org/10.1101/2020.03.11.20034363

9. Ciarochi J. How COVID-19 and other infectious diseases spread: Mathematical modeling. 2020; https://triplebyte.com/blog/modeling-infectious-diseases.

10. Elaydi S. An Introduction to Difference Equations. New York: Springer Science + Business Media, Inc.; 2005.

11. Livadiotis G., Assas L., Dennis B., Elaydi S., and Kwessi E. Kappa function as a unifying framework for discrete population modeling. Nat. Res. Mod. 2016; 29:130–144.

12. Hillman G. N., and Barrett A. D. T. Vaccinology: an essential guide. Chichester, West Sussex: Wiley Blackwell. 2015; p. 310.

13. Hao T. Infection dynamics of coronavirus disease 2019 (Covid-19) modeled with the integration of the Eyring’s rate process theory and free volume concept. MedRxiv. 2020; https://doi.org/10.1101/2020.02.26.20028571

14. Nishiura H. Time variations in the generation time of an infectious disease: implications for sampling to appropriately quantify transmission potential. Math. Biosci. Eng. 2010; 7: 851–869. https://doi.org/10.3934/mbe.2010.7.851 PMID: 21077712

15. Chen T., Rui J., Wang Q. P., Zhao Z.-Y., Cui J.-A., and Yin L. A mathematical model for simulating the phase-based transmissibility of a novel coronavirus. Infect. Dis. Poverty. 2020; 9:24. https://doi.org/10.1186/s40249-020-00640-3 PMID: 32111262

16. Livadiotis G., and Elaydi S. General Allee effect in two-species population biology. J. Biol. Dyn. 2012; 6: 959–973.

17. Kwessi E., Elaydi S., Livadiotis G., and Dennis B. Nearly Exact Discretization of Single Species Population Models. Nat. Res. Mod. 2018; 31:e12167.

18. Dayeh M.A., Livadiotis G., and Elaydi S. A Discrete Mathematical Model for the Aggregation of β-Amyloid. Plos One. 2018; 13:e0196402.

19. Wrapp D., Wang N., Corbett K. S., Goldsmith J. A., Hsieh C.-L., Abiona O., et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science, 2020; 367: 1260–1263.

20. Mozurkewich M., and Benson S. W. Negative activation energies and curved Arrhenius plots. 1. Theory of reactions over potential wells. J. Phys. Chem. 1984; 8:6429–6435.

21. Fasano G., and Vio R. Fitting a straight line with errors on both coordinates. Bulletin d'Information du Centre de Donnees Stellaires. 1988; 35:191–196.

22. Livadiotis G. Approach to general methods for fitting and their sensitivity. Physica A. 2007; 375:518–536.

23. Livadiotis G. Geometric interpretation of errors in multi-parametrical fitting methods. Stats. 2019; 2:426–438.

24. Frisch P.C., Bzowski M., Livadiotis G., McComas D.J., Mӧbius E., Mueller H.-R., et al. Decades-long changes of the interstellar wind through our solar system. Science. 2013; 341:1080.

25. Schwadron N.A., Mobius E., Kucharek H., Lee M.A., French J., Saul L., et al. Solar radiation pressure and local interstellar medium flow parameters from IBEX low energy hydrogen measurements. Astrophys. J. 2013; 775:86.

26. Fuselier S. A., Allegrini F., Bzowski M., Dayeh M. A., Desai M., Funsten H. O., et al. Low energy neutral atoms from the heliosheath. Astrophys. J. 2014; 784:89.
27. Livadiotis G. Chi-p distribution: Characterization of the goodness of the fitting using $L^p$ norms. J. Stat. Distr. Appl. 2014; 1:4.

28. Livadiotis G. General fitting methods based on $L_q$ norms and their optimization. Stats. 2020; 3:16–31.

29. Livadiotis G., and Desai M.I. Plasma-field coupling at small length scales in solar wind near 1 au. Astrophys. J. 2016; 829:88.