# Temperature dependence of COVID-19 transmission

Alessio Notari[1]*
[1] *Departament de Física Quàntica i Astrofisíca & Institut de Ciències del Cosmos (ICCUB), Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain*

The recent coronavirus pandemic follows in its early stages an almost exponential expansion, with the number of cases as a function of time reasonably well fit by $N(t) \propto e^{\alpha t}$, in many countries. We analyze the rate $\alpha$ in different countries, choosing as a starting point in each country the first day with 30 cases and fitting for the following 12 days, capturing thus the early exponential growth in a rather homogeneous way. We look for a link between the rate $\alpha$ and the average temperature $T$ of each country, in the month of the epidemic growth. We analyze a *base* set of 42 countries, which developed the epidemic earlier, and an *extended* set of 88 countries, which developed the epidemic more recently. Fitting with a linear behavior $\alpha(T)$, we find evidence in both datasets for a decreasing growth rate as a function of $T$, at 99.66%C.L. and 99.86%C.L. in the *base* and *extended* dataset, respectively. In the *base* set, going beyond a linear model, a peak at about $(7.7 \pm 3.6)^{\circ}C$ seems to be present in the data. Our findings give hope that, for northern hemisphere countries, the growth rate should significantly decrease as a result of both warmer weather and lockdown policies. In general the propagation should be hopefully stopped by strong lockdown, testing and tracking policies, before the arrival of the next cold season.

## I. INTRODUCTION

The recent coronavirus (COVID-19) pandemic is having a major effect in many countries, which needs to be faced with the highest degree of scrutiny. An important piece of information is whether the growth rate of the confirmed cases among the population could decrease with increasing temperature. Experimental research on related viruses found indeed a decrease at high temperature and humidity [1]. We try to address this question using available epidemiological data. A similar analysis for the data from January 20 to February 4, 2020, among 403 different Chinese cities, was performed in [2] and similar studies were recently performed in [3–7]. The paper is organized as follows. In section II we explain our methods, in section III we show the results of our analysis and in section IV we draw our conclusions.

## II. METHOD

We start our analysis from the empirical observation that the data for the coronavirus disease in many different countries follow a common pattern: once the number of confirmed cases reaches order 10 there is a very rapid subsequent growth, which is well fit by an exponential behavior. The latter is typically a good approximation for the following couple of weeks and, after this stage of *free* propagation, the exponential growth typically gradually slows down, probably due to other effects, such as: lockdown policies from governments, a higher degree of awareness in the population or the tracking and isolation of the positive cases.

Our aim is to see whether the temperature of the environment has an effect, and for this purpose we choose to analyze the first stage of *free* propagation in a selected sample of countries. We choose our sample using the following rules:

- we start analyzing data from the first day in which the number of cases in a given country reaches a reference number $N_i$, which we choose to be $N_i = 30$ [8];
- we include only countries with at least 12 days of data, after this starting point.

*Electronic address: notari@fqa.ub.edu

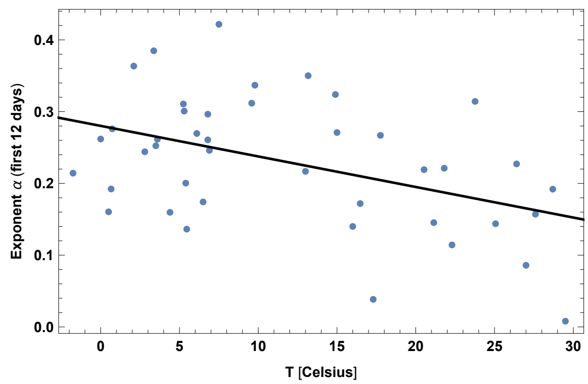


Figure 1: Exponent $\alpha$ for each country vs. average temperature $T$, for the relevant period of time, as defined in the text, for the base set of 42 countries. We show the data points and the best-fit for the linear interpolation.

The data were collected from [9]. We then fit the data for each country with a simple exponential curve $N(t) = N_0\, e^{\alpha t}$, with 2 parameters, $N_0$ and $\alpha$; here $t$ is in units of days. In the fit we used Poissonian errors, given by $\sqrt{N}$, on the daily counting of cases. We associated then to each country an average temperature $T$, for the relevant weeks, which we took from [10]. More precisely: if for a given country the average $T$ is tabulated only for its capital city, we directly used such a value. If, instead, more cities are present for a given country, we used an average of the temperatures of the main cities, weighted by their population [11]. For most countries we used the average temperature for the month of March, with a few exceptions [12].

We analyzed two datasets. A first list of countries was selected on March 26th. The list of such *base* dataset includes 42 countries: Argentina, Australia, Belgium, Brazil, Canada, Chile, China, Czech Republic, Denmark, Egypt, Finland, France, Germany, Greece, Iceland, India, Indonesia, Iran, Ireland, Israel, Italy, Lebanon, Japan, Malaysia, Netherlands, Norway, Philippines, Poland, Portugal, Romania, Saudi Arabia, Singapore, Slovenia, South Korea, Spain, Sweden, Switzerland, Taiwan, Thailand, United Arab Emirates, United Kingdom, U.S.A..

An additional set of countries was added to the first dataset on April 1st, reaching a total of 88 countries. The added countries are: Albania, Andorra, Algeria, Armenia, Austria, Bahrain, Bosnia and Herzegovina, Brunei, Bulgaria, Burkina Faso, Cambodia, Colombia, Costa Rica, Croatia, Cyprus, Dominican Republic, Ecuador, Estonia, Hungary, Iraq, Jordan, Kazakhstan, Kuwait, Latvia, Lithuania, Luxembourg, Malta, Mexico, Moldova, Morocco, New Zealand, North Macedonia, Oman, Panama, Pakistan, Peru, Qatar, Russia, Senegal, Serbia, Slovakia, South Africa, Tunisia, Turkey, Ukraine, Uruguay, Vietnam.

Using such datasets for $\alpha$ and $T$ for each country, we fit with two functions $\alpha(T)$, as explained in the next section. Note that the statistical errors on the $\alpha$ parameters, considering Poissonian errors on the daily counting of cases, are typically much smaller than the spread of the values of $\alpha$ among the various countries. This is due to systematic effects, which are dominant, as we will discuss later on. For this reason we disregarded statistical errors on $\alpha$. The analysis was done using the software *Mathematica*, from Wolfram Research, Inc..

## III. RESULTS

We first fit the *base* dataset, with a simple linear function $\alpha(T) = \alpha_0 + \beta\, T$, to look for an overall decreasing behavior. Results for the best fit, together with our data points, are shown in fig. 1. The parameter estimation contours are shown in fig. 2, and the estimate, standard deviation and confidence intervals for the parameters are shown in Table I. From such results a clear decreasing trend is visible, and indeed the slope $\beta$ is negative, at 99.66% C.L. ($p$-value 0.0034).

| Parameter | estimate | $\sigma$ | 95% lower | 95% upper |
|---|---|---|---|---|
| $\alpha_0$ | 0.280 | 0.021 | 0.238 | 0.321 |
| $\beta$ | −0.00425 | 0.00136 | −0.00701 | −0.00149 |

Table I: Best-estimate, standard deviation ($\sigma$) and 95% C.L. intervals for the parameters of the linear interpolation, for the base set of 42 countries.

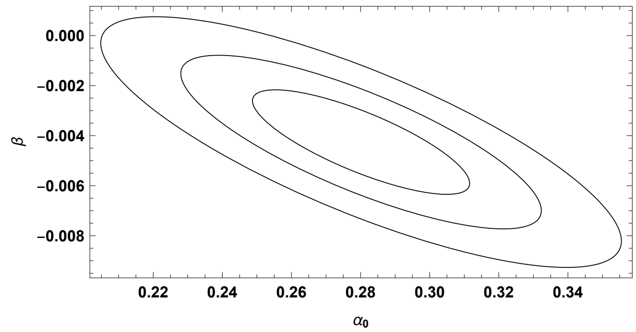


Figure 2: Confidence regions for the parameters of the linear model, for the base set of 42 countries. Contours represent 68% C.L., 95% C.L. and 99.7% C.L. respectively.

However, the linear fit is able to explain only a small part of the variance of the data, with $R^2 = 0.196$, and its adjusted value $R^2_{\text{adjusted}} = 0.175$, clearly due to the presence of many more factors.

In addition, a decreasing trend is also visible in this dataset, below about $10°C$. For this reason we also fit with a quadratic function $\alpha(T) = \alpha_0 - \beta(T - T_M)^2$. Results for the quadratic best fit are presented in fig. 3. The relative estimate, standard deviation and confidence intervals for the parameters are shown in Table II. From such results a peak is visible at around $T_M \approx 8°C$. The quadratic model is able to explain a slightly larger part of the variance of the data, since $R^2 \approx 0.27$ [13]. Moreover, despite the presence of an extra parameter, one may quantify the improvement of the fit, using for instance the Akaike Information Criterion (AIC) for model comparison, $\Delta\text{AIC} \equiv 2\Delta k - 2\Delta \ln(\mathcal{L})$, where $\Delta k$ is the increase in the number of parameters, compared to the simple linear model, and $\Delta \ln(\mathcal{L})$ is the change in the maximum log-likelihood between the two models. This gives $\Delta\text{AIC} = -2.1$, slightly in favor of the quadratic model.

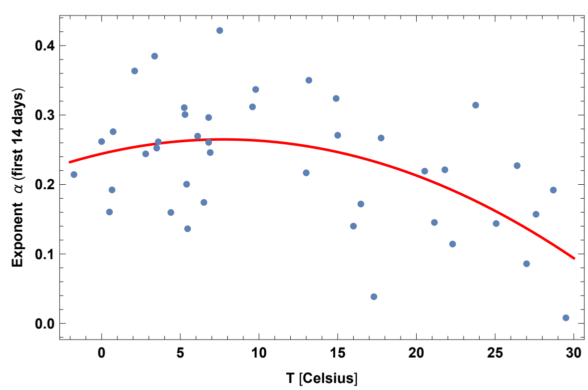


Figure 3: Exponent $\alpha$ for each country vs. average temperature $T$, as defined in the text, for the base set of 42 countries. We show here the quadratic best-fit.

| Parameter | estimate | $\sigma$ | 95% lower | 95% upper |
|---|---|---|---|---|
| $\alpha_0$ | 0.264 | 0.0159 | 0.2325 | 0.2972 |
| $\beta$ | 0.000345 | 0.000173 | $-5.104 \cdot 10^{-6}$ | 0.000694 |
| $T_M$ | 7.73 | 3.64 | 0.37 | 15.1 |

Table II: Best-estimate, standard deviation ($\sigma$) and 95% C.L. intervals for the parameters of the quadratic interpolation, for the base set of 42 countries.

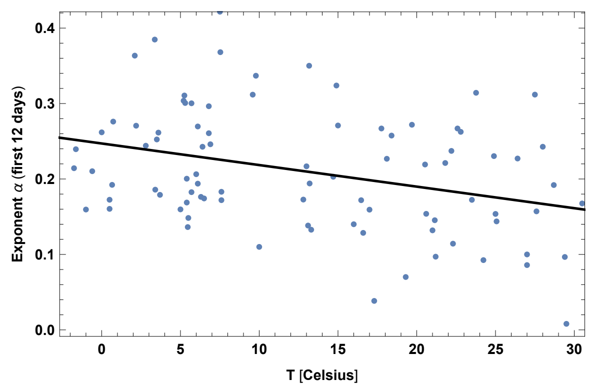


Figure 4: Exponent $\alpha$ for each country vs. average temperature $T$, for the relevant period of time, as defined in the text, for the extended set of 88 countries. We show the data points and the best-fit for the linear interpolation.

We repeat then the same analysis for the *extended* dataset of 88 countries. Results for the linear fit, together with our data points, are shown in fig. 4. The parameter estimation contours are shown in fig. 5, and the estimate, standard deviation and confidence intervals for the parameters are shown in Table III. The slope $\beta$ is smaller in absolute value, but the significance remains high, since a zero slope is excluded at 99.86% C.L. ($p$-value 0.0014). Now $R^2 = 0.11$ and $R^2_{\text{adjusted}} = 0.10$.

In this enlarged sample the quadratic trend is not visible anymore, and indeed the AIC does not prefer the quadratic fit: $\Delta\text{AIC} = +0.9$ compared to the linear fit, in disfavor of the quadratic model. The $R^2$ is also practically the same as in the linear fit.

| Parameter | estimate | $\sigma$ | 95% lower | 95% upper |
|---|---|---|---|---|
| $\alpha_0$ | 0.247 | 0.0138 | 0.220 | 0.275 |
| $\beta$ | $-0.00286$ | 0.000867 | $-0.00458$ | $-0.00113$ |

Table III: Best-estimate, standard deviation ($\sigma$) and 95% C.L. intervals for the parameters of the linear interpolation, for the extended set of 88 countries.

## IV. DISCUSSION AND CONCLUSIONS

We have collected data for countries that had at least 12 days of data after a starting point, which we fixed to be at the threshold of 30 confirmed cases. We considered two datates: a *base* dataset with 42 countries, collected on March 26th, and an *extended* dataset with a total of 88 countries, collected on April 1st. We have fit the data for each country with an exponential and extracted the exponents $\alpha$, for each country. Then we have analyzed such exponents as a function of the temperature $T$, using the average temperature for the month of March (or slightly earlier in some cases), for each of the selected countries.

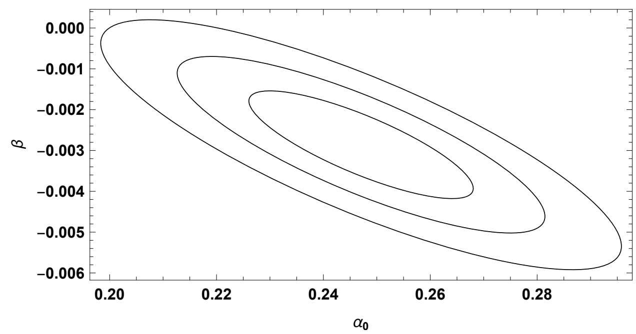


Figure 5: Confidence regions for the parameters of the linear model, for the extended set of 88 countries. Contours represent 68% C.L., 95% C.L. and 99.7% C.L., respectively.

For the *base* dataset we have shown that the growth rate of the transmission of the COVID-19 has a decreasing trend, as a function of $T$, at 99.66% C.L. ($p$-value 0.0034). In this fit $R^2 = 0.196$. In addition, using a quadratic fit, we have shown that a peak of maximal transmission seems to be present in this dataset at around $(7.7 \pm 3.6)°C$. Such findings are in good agreement with a similar study, performed for Chinese cities [2], which also finds the existence of an analoguous peak and an overall decreasing trend. Other similar recent studies [3–6] find results which seem to be also in qualitative agreement.

For the extended dataset we also found a decreasing slope $\beta$. This is smaller in absolute value, but the significance remains high, since a zero slope is excluded at 99.86% C.L. ($p$-value 0.0014). For this fit we found $R^2 = 0.11$.

The decrease at high temperatures is expected, since the same happens also for other coronaviruses [1]. It is unclear instead how to interpret the decrease at low temperature (less than $8°C$), present in the *base* dataset. This could be a statistical fluctuation, since it is not present in the *extended* dataset. One possible reason for this decrease, if real, could be the lower degree of interaction among people in countries with very low temperatures, which could slow down the propagation of the virus.

A general observation is also that a large scatter in the residual data is present, clearly due to many other systematic factors, such as variations in the methods and resources used for collecting data and variations in the amount of social interactions, due to cultural reasons. It is also possible that variations in resources bias the testing procedure (*i.e.* poorer countries have less intense testing), which might be partially degenerate with effects of temperature. Further study would be required to assess such factors.

As a final remark, our findings can be very useful for policy makers, since they support the expectation that with growing temperatures the coronavirus crisis should become milder in the coming few months, for countries in the Northern Hemisphere. As an example the estimated doubling time, with the quadratic fit, at the peak temperature of $7.7°C$ is of 2.6 days, while at $26°C$ is expected to go to about 4.6 days. The linear fit gives a smaller effect: a doubling time of 2.8 days (or 3.1 days) at $7.7°C$ and a doubling time of 4.1 days (or 4 days) at $26°C$, using the estimate from the *base* (or the *extended*) dataset. For countries with seasonal variations in the Southern Hemisphere, instead, this should give motivation to implement strong lockdown policies before the arrival of the cold season.

We stress that, in general, it is important to fully stop the propagation, using strong lockdown, testing and tracking policies, taking also advantage of the warmer season, and before the arrival of the next cold season.

**Acknowledgments**

We would like to acknowledge Viviana Acquaviva, Alberto Belloni, Ángel J. Gómez Peláez, Jordi Miralda and Giorgio Torrieri, for useful discussions and comments.

---

[1] K. H. Chan, J. S. Malik Peiris, et al. *The Effects of Temperature and Relative Humidity on the Viability of the SARS Coronavirus. Advances in Virology.*, Vol. 2011(Article ID 734690), 2011.
[2] Mao Wang, Aili Jiang et al. *Temperature significant change COVID-19 Transmission in 429 cities.* 2020. https://www.medrxiv.org/content/early/2020/02/25/2020.02.22.20025791.
[3] Miguel B. Araujo and Babak Naimi. *Spread of SARS-CoV-2 Coronavirus likely to be constrained by climate.* https://www.medrxiv.org/content/early/2020/03/16/2020.03.12.20034728. 2020.
[4] Qasim Bukhari and Yusuf Jameel. *Will Coronavirus Pandemic Diminish by Summer?* ; available at ssrn: https://ssrn.com/abstract=3556998. 2020.
[5] Kai Feng Weifeng Lv Jingyuan Wang, Ke Tang. *High Temperature and High Humidity Reduce the Transmission of COVID-19.* 2020. arxiv:2003.05003 [q-bio.pe]. 2020.
[6] Mohammad M. Sajadi, Parham Habibzadeh, Augustin Vintzileos, Shervin Shokouhi, Fernando Miralles-Wilhelm, and Anthony Amoroso. *Temperature, Humidity and Latitude Analysis to Predict Potential Spread and Seasonality for COVID-19* ; available at ssrn: https://ssrn.com/abstract=3550308 or http://dx.doi.org/10.2139/ssrn.3550308.
[7] Wei Luo, Maimuna S Majumder, Dianbo Liu, Canelle Poirier, Kenneth D Mandl, Marc Lipsitch, and Mauricio Santillana. *The role of absolute humidity on transmission rates of the COVID-19 outbreak.* https://www.medrxiv.org/content/early/2020/02/17/2020.02.12.20022467. 2020.
[8] In practice we choose, as the first day, the one in which the number of cases $N_i$ is closest to 30. In some countries, such a number $N_i$ is repeated for several days; in such cases we choose the last of such days as the starting point.
[9] https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases
[10] https://en.wikipedia.org/wiki/List_of_cities_by_average_temperature
[11] The only two exceptions to this procedure are: Japan and U.S.A.. For Japan we have subdivided into three regions: Hokkaido, Okinawa and the rest of the country, using respectively the temperatures of Sapporo, Naha and Tokio. For the U.S.A. we used the national average of about 5.3 degrees from https://www.ncdc.noaa.gov/sotc/national/201903. For Ecuador, we used the average $T = 27.5°C$ of Guayaquil, the main site for the disease. For Brunei we used $T = 27\,°C$ and for Morocco we used the average Temperature for Casablanca, $T = 14.7°C$, from https://en.climate-data.org. For Jordan we used $T = 17°C$, from https://www.weather-atlas.com/en/jordan/amman-weather-march.
[12] For China, South Korea, Singapore, Iran, Taiwan and Japan we considered an interpolating function of the temperature for the months of January, February and March and we took an average of such function in the relevant 12 days of the epidemic.
[13] Here $R^2$ is defined as $R^2 \equiv 1 - \frac{SS_R}{SS_T}$, where $SS_R$ is the residual sum of squares and $SS_T$ is the sum of the squared differences between the $\alpha$ values and their mean value.