Science of the Total Environment 763 (2021) 144390

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Temperature dependence of COVID-19 transmission

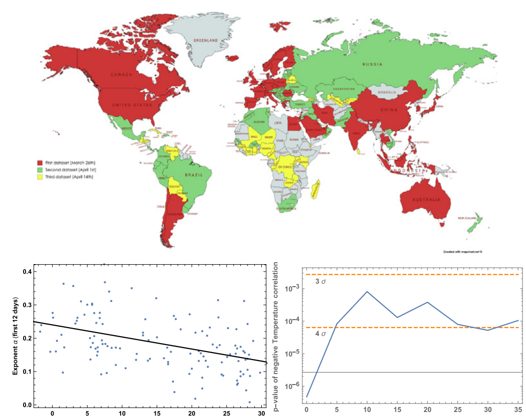
Alessio Notari

Departament de Física Quàntica i Astrofisíca, Institut de Ciències del Cosmos (ICCUB), Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain

HIGHLIGHTS

- We analyze the initial spread of COVID cases for each country with an exponential fit.
- We use 12 days of data after 30 cases, for 3 datasets: with 42, 88 and 125 countries.
- COVID-19 spread is slower at high Temperature in all datasets with high significance.
- We use GDP per capita (GDPPC) as an indicator of testing capability.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 7 September 2020
Received in revised form 4 December 2020
Accepted 4 December 2020
Available online 13 December 2020

Editor: Jay Gan

Keywords:
COVID-19
Environmental risk factors
Epidemiology

ABSTRACT

The recent COVID-19 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 α in different countries, starting in each country from a threshold of 30 total 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 α and the average temperature T of each country, in the month of the initial epidemic growth. We analyze a *base* set of 42 countries, which developed the epidemic at an earlier stage, an *intermediate* set of 88 countries and an *extended* set of 125 countries, which developed the epidemic more recently. Fitting with a linear behavior $\alpha(T)$, we find increasing evidence in the three datasets for a slower spread at high T, at 99.66% C.L., 99.86% C.L. and 99.99995% C.L. (p-value $5 \cdot 10^{-7}$, or 5σ detection) in the *base*, *intermediate* and *extended* dataset, respectively. The doubling time at 25 °C is 40% ~ 50% longer than at 5 °C. Moreover we analyzed the possible existence of a bias: poor countries, typically located in warm regions, might have less intense testing. By excluding countries below a given GDP per capita from the dataset, we find that this affects our conclusions only slightly and only for the *extended* dataset. The significance always remains high, with a p-value of about 10^{-3} - 10^{-4} or less. 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, policy measures should be taken to prevent a second wave, such as safe ventilation in public buildings, social distancing, use of masks, testing and tracking policies, before the arrival of the next cold season.

1. Introduction

The recent coronavirus (COVID-19) pandemic is having a major effect in many countries, which needs to be faced with the highest degree

E-mail address: notari@fqa.ub.edu.

https://doi.org/10.1016/j.scitotenv.2020.144390
0048-9697/

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 (Chan et al., 2011). 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 Mao et al. (2020) and similar studies were recently performed in Araujo and Naimi (2020), Bukhari and Jameel (2020), Feng et al. (2020), Luo et al. (2020), and Sajadi et al. (n.d.). More studies have subsequently confirmed this conclusion, while other studies found mixed results or even in contradiction with it, as we will review in the Discussion section.

The paper is organized as follows. In Section 2 we explain our methods, in Section 3 we show the results of our analysis and in Section 5 we draw our conclusions.

2. 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 of infection from SARS-CoV-2 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 and/or the tracking and isolation of the positive cases.

Our aim is to see whether the temperature of the environment has a correlation with the disease spread, 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[1];
- we include only countries with at least 12 days of data, after this starting point.

The data were collected from.[2] 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 α; here t is in units of days. See e.g. Dehning et al. (2020) for a justification of the exponential behavior. We associated then to each country an average temperature T, for the relevant weeks, which we took from.[3] 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.[4] We used the average temperature in the relevant time range by interpolating monthly data (using the month of the epidemic together with the next and the previous month), constructing an interpolating curve $T(d)$ as a function of the day. Then, by using such an interpolation we used the value $T(d_i + 6)$, in the middle of the chosen range, with duration D=12 days.

We analyze in Section 3 how the parameter reconstruction varies when using different choices for the total number of days D, which will fully justify our choice of 12 days.

We analyzed three datasets, as shown in Fig. 1. The three datasets are simply different groups analyzed at different times of the disease spread in the planet. 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, Thailand, United Arab Emirates, United Kingdom, U.S.A.[5]

An *intermediate* dataset was built by adding 46 countries (the second dataset in Fig. 1, built on April 1st) to the first dataset, 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.

Finally an *extended* set was built on April 14th,[6] adding the following countries (third dataset in Fig. 1) to the previous dataset: Belarus, Bolivia, Cameroon, Congo, Cote d'Ivoire, Cuba, Democratic Republic of Congo, Djibouti, El Salvador, Georgia, Ghana, Guatemala, Guinea, Honduras, Jamaica, Kenya, Kosovo, Kyrgyzstan, Madagascar, Mali, Mauritius, Montenegro, Niger, Nigeria, Paraguay, Puerto Rico, Rwanda, Sri Lanka, Togo, Trinidad and Tobago, Uganda, Uzbekistan, Venezuela, Zambia.

Using such datasets for α 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 α parameters, considering Poissonian errors on the daily counting of cases, are typically much smaller than the spread of the values of α 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 α. The analysis was done using the software *Mathematica*, from Wolfram Research, Inc.

3. 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. 2. The estimate, standard deviation and confidence intervals for the parameters, together with the significance and the explained variance, R^2, are shown in Table 1. From such results a clear decreasing trend is visible, and indeed the slope β is negative, at 99.66% Confidence Level (C.L.), or equivalently with a p-value of 0.0034.

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_{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 and in Table 2. From such results a peak is visible at around $T_M \approx 8\ ^\circ C$. The quadratic model is able to explain a slightly larger part of the variance of the data, since $R^2 \approx 0.27$.[7] Moreover, despite the

[1] 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. For the particular case of China, we started from January 16th, with 59 cases, since the number before that day was essentially frozen.

[2] https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases.

[3] https://en.wikipedia.org/wiki/List_of_cities_by_average_temperature. For some cases data was missing and so we took it from https://en.climate-data.org and from https://www.weather-atlas.com/.

[4] The only two exceptions to this procedure are: Japan, U.S.A. and Ecuador. For Japan we 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.

[5] We also added Taiwan as a data point, since data are fully available, irrespective of its political status, which is irrelevant for our discussion.

[6] Only countries with at least 300.000 inhabitants were considered in this dataset.

[7] 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 SST is the sum of the squared differences between the α values and their mean value.

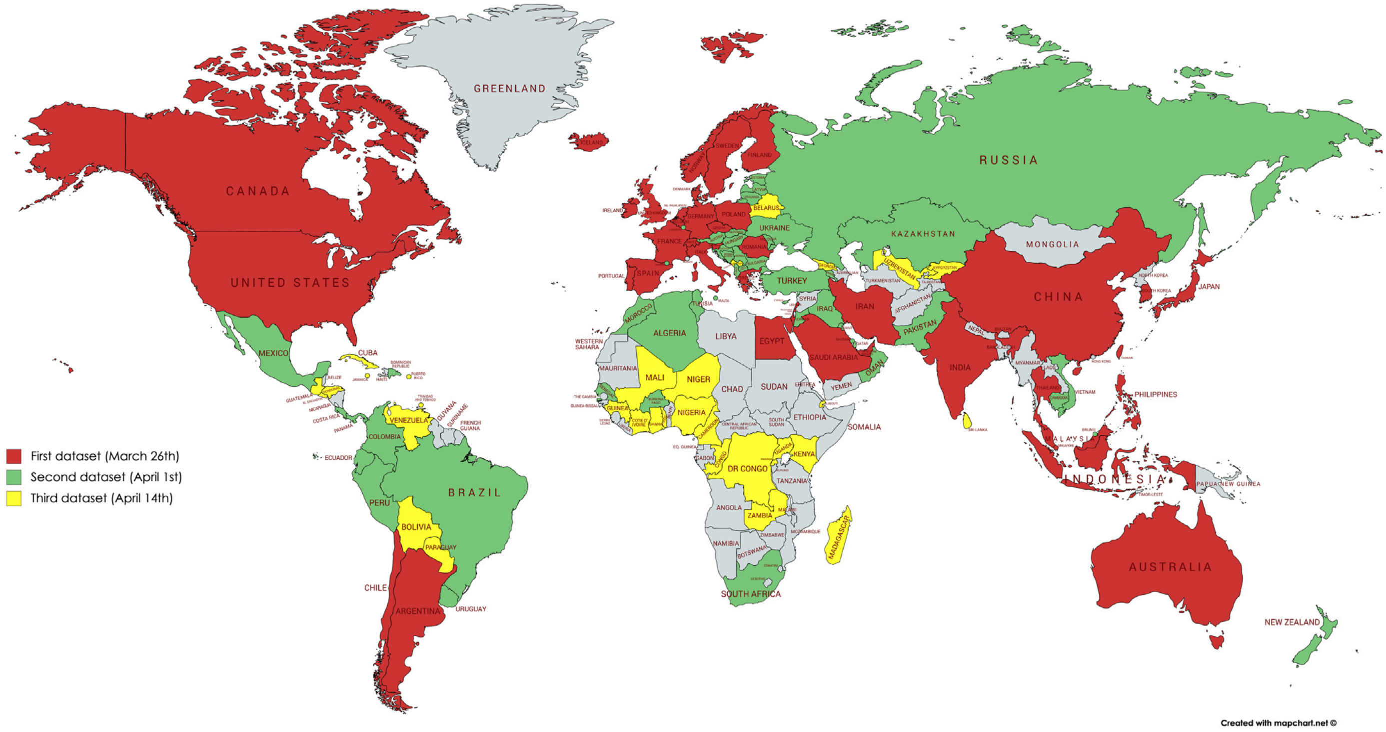


Fig. 1. Countries included in the analysis: the *base* dataset is in red, the *intermediate* dataset includes both countries in red and in green, and the *extended* dataset includes red, green and yellow.

presence of an extra parameter, one may quantify the improvement of the fit, using for instance the Akaike Information Criterion (AIC) for model comparison, $\Delta\text{AIC} \equiv 2\Delta k - 2\Delta \ln(\mathscr{L})$, where Δk is the increase in the number of parameters, compared to the simple linear model, and $\Delta \ln(\mathscr{L})$ 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.

We repeated then the same analysis for the *intermediate* dataset of 88 countries and for the *extended* dataset of 125 countries. Results for the linear fit of the *intermediate* sample are shown in Fig. 4 and in Table 3. The slope β is smaller in absolute value, but the significance actually slightly increases, since a zero slope is excluded at 99.86% C.L. (p-value 0.0014). Now R^2=0.11 and R^2_{adjusted}=0.10.

In this 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.

For the *extended* sample results of the linear fit are shown in Fig. 5 and in Table 4. The slope β becomes larger and, most importantly, the significance highly increases, since a zero slope is now excluded at 99.99995% C.L. (p-value $5\cdot10^{-7}$, or 5σ detection, translated in the language of a Gaussian distribution). Now R^2=0.19 and R^2_{adjusted}= 0.18.

In this dataset, which extends to April 14th, a few anomalies are however present: in the case of Bangladesh and Thailand it is possible to see that the exponential growth became much faster after the initial 12 days. We checked what happens by using a different interval of time for these 2 cases, instead of the standard 12 days. Namely, we used 44 days for Thailand and 21 days for Bangladesh, which give the maximal value of α in both cases. The results for the linear fits using such corrected values is shown in Table 5. The significance is lower, but still very high: p-value $4.6\cdot10^{-6}$, or 4.6σ detection, translated in the language of a Gaussian distribution.

We also tested the existence of a possible bias on the data: the fact that poor countries have less intense testing. This could in principle be a source of major bias, since many countries with low income are located in warm regions. In order to discard such a bias we analyzed the existence of a nonzero linear correlation β on subsamples of the *extended* dataset, by excluding countries with low

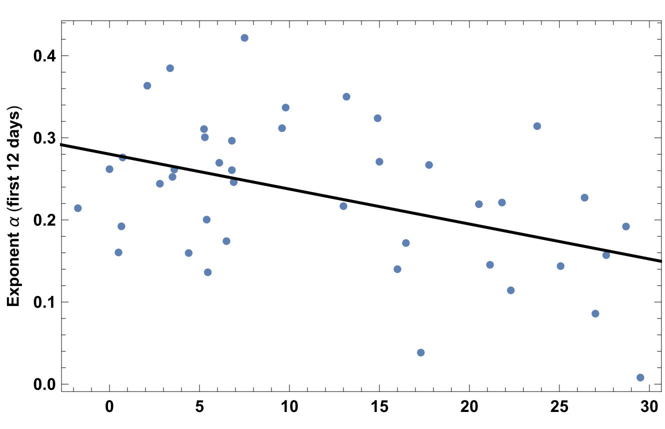


Fig. 2. Exponent α for each country vs. average temperature T, for the relevant period of time, as defined in the text, for the base set of 42 countries. We show the data points and the best-fit for the linear interpolation.

Table 1
In the left panel: best-estimate, standard deviation (σ) and 95% C.L. intervals for the parameters of the linear interpolation, for the *base* set of 42 countries. In the right panel: R^2 for the best-estimate and $p-$value for a non-zero β.

Parameter	Estimate	σ	95% lower	95% upper
α_0	0.280	0.021	0.238	0.321
β	−0.00425	0.00136	−0.00701	−0.00149
R^2				0.196
p-Value				0.0033

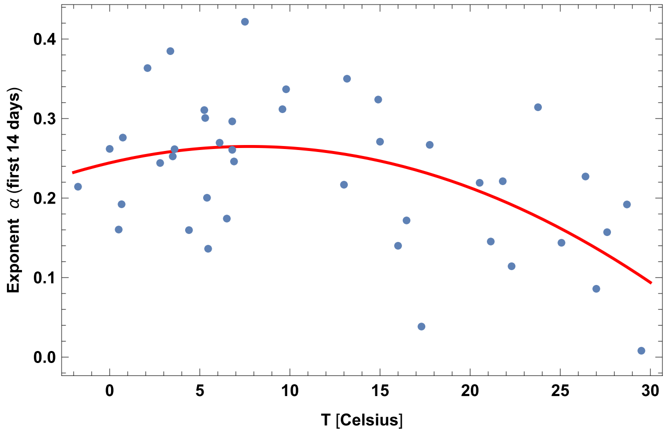


Fig. 3. Exponent α for each country vs. average temperature T, as defined in the text, for the base set of 42 countries. We show here the quadratic best-fit.

income. More specifically we set a threshold on the GDP per capita,[8] and checked whether the correlation is still there, excluding countries below such a threshold from the analysis. We show in Fig. 6 our results: we find a correlation to exist, rather independently on the threshold that we applied. The significance of a nonzero beta (p-value) is plotted in Fig. 7 and remains always between $5 \cdot 10^{-7}$ and $8 \cdot 10^{-4}$.

In addition, we also checked for a correlation between the growth rate α and the GDP per capita, shortly *GDP*. We find *no* significant correlation in the *base* and *intermediate* datasets, while we find a negative correlation in the *extended* dataset, with p-value $=0.0012$. This is not so surprising, since the *extended* dataset contains many low-income countries, where the disease has arrived later, and where most likely testing is not intense enough. For this dataset we performed thus a linear fit with two variables, *GDP* and *T*. Results are shown in Table 6. The dependence on T is still highly significant, with p-value $\simeq 0.000048$ and the best-estimate is $\beta \simeq -0.0031$. As expected, T also has a non-negligible correlation with the GDP per capita.

Interestingly, when considering only countries above 5 thousand dollars GDP per capita in the *extended* dataset, the correlation of α with *GDP* per capita becomes insignificant. This should be probably interpreted as the fact that testing capabilities do not induce a bias as long as the GDP per capita is not very low.

Finally we performed some tests, by varying the number of days D used in our analysis. First, we tested the behavior of the statistical errors on the parameter α extracted from the exponential fits. Each country has a best estimate for α and an error $\delta\alpha$: we show in Fig. 8 the mean error $\langle\delta\alpha\rangle$ over the full sample of 126 countries, finding that it decreases with increasing D, as it should. Regarding the parameter N_0 we find for $D=12$ that the mean is $\langle N_0\rangle = 33.8 \pm 4.1$, which is fully consistent with the choice $N_i=30$ for the first day ($t=0$).

Then we also repeated the correlation analysis with temperature for different values of D, in order to test whether the choice $D=12$ is optimal. We show in Fig. 9 the dependence of the linear correlation coefficient β with D. As expected, the error σ_β on the correlation β decreases with increasing D simply because of using more data. For very short durations (D less than 4 days) there is no significant detection of a nonzero beta. On the other hand, for large durations D, even if the statistical errors become smaller, the Temperature effect is expected to become more difficult to disentangle from other effects (such as lockdowns, social distancing and other measures from the governments, which can take place if D is large). Indeed the significance of the Temperature effect (represented by the t-statistic β/σ_β in Fig. 9) is

Table 2
In the left panel: best-estimate, standard deviation (σ) and 95% C.L. intervals for the parameters of the quadratic interpolation, for the *base* set of 42 countries. In the right panel: R^2 for the best-estimate and p−value for a non-zero β.

Parameter	Estimate	σ	95% lower	95% upper
α_0	0.264	0.0159	0.2325	0.2972
β	0.000345	0.000173	$-5.104 \cdot 10^{-6}$	0.000694
T_M	7.73	3.64	0.37	15.1
R^2				0.27
p-Value (β)				0.053

maximal for a duration D between 12 and 20 days, which fully justifies our choice of 12 days.

4. Discussion

After the publication of the initial preprints of the present work, several authors analyzed the correlation of the disease spread with Temperature, finding a similar correlation worldwide (Huang et al., 2020; Liu et al., n.d.; Mandal and Panwar, 2020; Ozyigit, 2020; Wu et al., 2020). Other studies focused on particular regions: in Brazilian cities a negative correlation was found up to about 25 °C (Prata et al., 2020), in China a negative correlation was found in He et al. (2020) and Shi et al. (2020) (while a positive correlation was found below 3 °C (Xie and Zhu, 2020)). Significant negative correlations were found in several Latin american cities (Bolano-Ortiz et al., 2020), in the U.S.A. (Adam et al., 2020) and in Japan (Mugen et al., 2020). The fact that many studies found similar effects with different datasets, also at the regional level, is a further confirmation of the significance of the correlation with Temperature.

Other studies found mixed results, in very specific places: a positive correlation in Singapore (Pani et al., 2020) and in Oslo (Menebo, 2020), and both positive and negative correlations for some Chinese provinces (Shahzad et al., 2020).

Correlations in agreement with ours, with both GDP and temperature were found also in Sarmadi et al. (2020).

In Jüni et al. (2020), instead, absence of correlation with Temperature and GDP was claimed, using the ratio of the number of cases in two fixed days in many different geographical areas. Such a study, however, seems to have methodological problems since it compares data in different areas at the same time, i.e. at very different stages of the contagion; this introduces large biases, since it is clear that if the epidemic has been going on for several weeks the effects of lockdowns and government policies dominate the spread of the disease. Furthermore, instead of using all the data of the curves, Jüni et al. (2020) used only the ratio

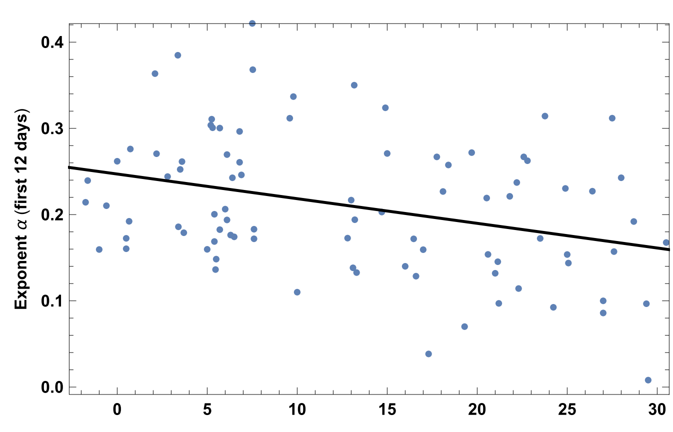


Fig. 4. Exponent α for each country vs. average temperature T, for the relevant period of time, as defined in the text, for the *intermediate* set of 88 countries. We show the data points and the best-fit for the linear interpolation.

[8] We used here data from https://ourworldindata.org/ on real GDP per capita, for the year 2017.

Table 3
In the left panel: best-estimate, standard deviation (σ) and 95% C.L. intervals for the parameters of the linear interpolation, for the *intermediate* set of 88 countries. In the right panel: R^2 for the best-estimate and $p-$value for a non-zero β.

Parameter	Estimate	σ	95% lower	95% upper
α_0	0.247	0.0138	0.220	0.275
β	−0.00286	0.000867	−0.00458	−0.00113
R^2				0.11
p-Value				0.0014

Table 4
In the left panel: best-estimate, standard deviation (σ) and 95% C.L. intervals for the parameters of the linear interpolation, for the *extended* set of 125 countries. In the right panel: R^2 for the best-estimate and $p-$value for a non-zero β.

Parameter	Estimate	σ	95% lower	95% upper
α_0	0.239	0.0125	0.215	0.264
β	−0.00357	0.000678	−0.00491	−0.00223
R^2				0.18
p-Value				$5.7 \cdot 10^{-7}$

between the initial day and the final day: this introduces very large measurement errors, due to large statistical fluctuations at these two specific days.

Note also that we are not claiming here that temperature is the only factor nor the dominant factor: indeed there must be other factors, since Temperature explains only about 18% of the variability of the disease spread (R^2=0.18). We have indeed found many other factors in the companion paper (Notari and Torrieri, 2020). Note also that some authors (Baker et al., 2020) have stressed that immunity among the population is more important than seasonal variations. This however is likely to be irrelevant for our analysis, since in the first days of the epidemic growth it should play no role. The same applies to lockdown policies: clearly they have a major impact, since they are able to cut the exponential growth, but they only do so in the long run. Moreover there could also be important non-linear interactions in the long run among immunity (Bansal et al., n.d.), lockdown policies and seasonality, which is an interesting point for future work, that goes beyond a linear model.

5. Conclusions

We 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 three datasets: a *base* dataset with 42 countries, collected on March 26th, an *intermediate* dataset with a total of 88 countries, collected on April 1st, and an *extended* dataset with a total of 125 countries, collected on April 14th. We fitted the data for each country with an exponential and extracted the exponents α. Then we performed a fit of 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.

For the *base* dataset we showed that the growth rate of the COVID-19 transmission 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 showed that a peak of maximal transmission seems to be present in this dataset at around $(7.7 \pm 3.6)\,^\circ C$. Such findings are in good agreement with a similar study, performed for Chinese cities (Mao et al., 2020), which also finds the existence of an analogous peak and an overall decreasing trend. Other similar previous studies (Araujo and Naimi, 2020; Bukhari and Jameel, 2020; Feng et al., 2020; Sajadi et al., n.d.) found results which seem to be also in qualitative agreement. Many other studies, subsequent to the preprints of our work, found similar results, as we reviewed in the Discussion section.

For the *intermediate* dataset we also found a decreasing slope β. 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.

Finally for the *extended* dataset we found a very highly significance for a negative β, p-value $5 \cdot 10^{-6} \sim 5 \cdot 10^{-7}$ (depending on the treatment of some anomalous cases), which would translate in a $4.5\sigma \sim 5\sigma$ detection, in the language of Gaussian distributions. Here R^2=0.16 ~ 0.2.

For all datasets we also tested the influence of a possible large bias: the fact that poorer countries have less intense testing, which might be in principle partially degenerate with effects of temperature. Our analysis indicate that this should not be a major issue: by excluding countries with low income from the analysis we find small variations on the best-fit value of β, and the significance of the correlation β remains very high, with p-value $8 \cdot 10^{-4}$ or less. We also checked for a correlation between the GDP per capita and α: we find a significant correlation only in the *extended* dataset. However, after taking into account of this variable, the dependence on T remains highly significant. Moreover, when considering only countries above 5 thousand dollars GDP per capita, the correlation of α with GDP per capita becomes insignificant. This should be probably interpreted as the fact that testing capabilities do not induce a bias as long as the GDP per capita is not too low.

The decrease at high temperatures is expected, since the same happens also for other coronaviruses (Chan et al., 2011). 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 *intermediate* and *extended* datasets. 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.

It is possible to make some hypothesis on possible causal explanations of our findings: (1) less resistance of the virus in aerosols due to UV radiation and higher temperature (Ratnesar-Shumate et al., 2020),

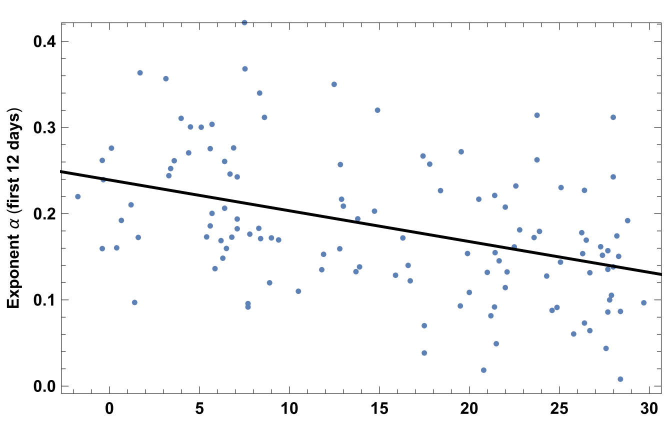


Fig. 5. Exponent α for each country vs. average temperature T, for the relevant period of time, as defined in the text, for the *extended* set of 125 countries. We show the data points and the best-fit for the linear interpolation.

Table 5
In the left panel: best-estimate, standard deviation (σ) and 95% C.L. intervals for the parameters of the linear interpolation, for the *extended* set of 125 countries. Here Thailand and Bangladesh were corrected for, as explained in the text. In the right panel: R^2 for the best-estimate and $p-$value for a non-zero β.

Parameter	Estimate	σ	95% lower	95% upper
α_0	0.2364	0.01235	0.2120	0.2609
β	−0.00321	0.0006699	−0.004538	−0.001885
R^2				0.16
p-Value				$4.6 \cdot 10^{-6}$

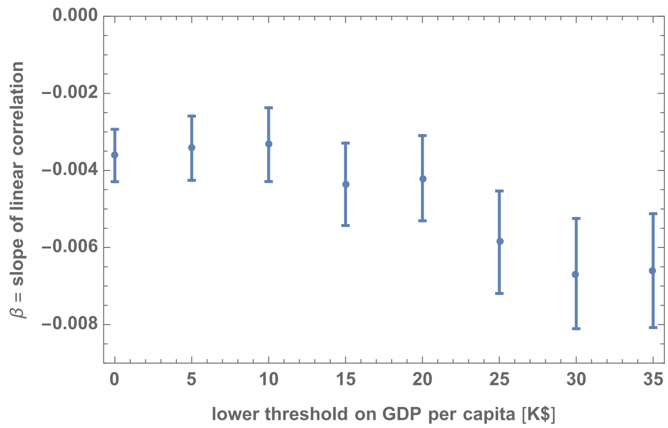


Fig. 6. We show the best estimate and the standard deviation for the parameter β of the linear model, excluding countries with a GDP per capita below a given threshold (in units of thousand dollars) from the *extended* set of 125 countries.

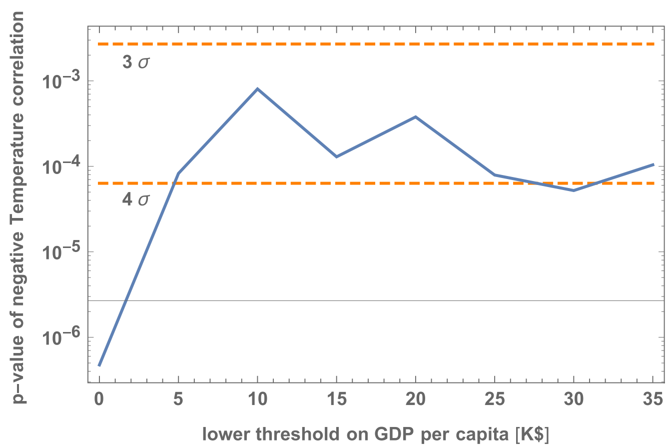


Fig. 7. We show the significance (p-value) for a nonzero parameter β, excluding countries with a GDP per capita below a given threshold in units of thousand dollars, from the *extended* set of 125 countries.

(2) better functioning of the immune system at higher temperature,[9] (3) at high temperature it is less likely to have a large number of people indoors in the same environment for extended periods of time. And, of course, a combination of all such factors is likely.

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. In a companion paper (Notari and Torrieri, 2020), we studied correlations with many other variables and, even after taking those variables into account, Temperature always remains a significant factor.

Some other limitations are present in our study. First, we considered GDP per capita to be an indicator of the testing capabilities of a given country. This could be improved by using more direct indicators, such as the number of available nasal swabs for COVID-19 or the health-care expenditure or the number of hospitals per capita.

However, the fact that many studies found similar correlations with Temperature with different datasets, also at the regional level, is a further confirmation that lack of testing should not be a

Table 6
In the top panel: best-estimate, standard error (σ), $t-$statistic and $p-$value for the parameters of the linear fit in two-variables, temperature (T) and GDP per capita (*GDP*), for the *extended* set of 125 countries. In the bottom panel: R^2 and correlation coefficient (i.e. normalized off-diagonal element of the covariance matrix) between *T* and *GDP*.

	Estimate	Standard error	t-Statistic	*p*-Value
1	0.2186	0.01795	12.17	8.15^{-23}
GDP	$6.165\cdot10^{-7}$	$3.78\cdot10^{-7}$	1.627	0.1061
T	−0.003118	0.0007397	−4.215	0.000048
R^2				0.2
$T-$ *GDP* correlation				0.41

major issue. Indeed, at regional level, the possible lack of testing capabilities should not represent a significant bias, since testing capabilities vary much less within a country than among different countries.

A second limitation of our study is the use of average Temperature, which is not very accurate for large countries, especially those that have a large spread in latitude and in climatic conditions, such as China or the U.S. or Russia. Possible approaches in order to improve on the present analysis could be: (i) to exclude very large countries from the data; (2) to split large countries into regions; (3) to add to Temperature a statistical error, based on the spread within each country. Clearly a city based approach could also improve the analysis, but this would require the existence of a worldwide database for COVID-19 spread for all major cities in the world.

Finally other environmental factors, such as humidity, wind speed, air pressure and pollution could also possibly play a role.

Some of the above points are analyzed in Notari and Torrieri (2020), while other ones are postponed to future work.

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 summer, for countries in the Northern Hemisphere. As an example the estimated doubling time, with the quadratic fit, at the peak temperature of 7.7 $^\circ C$ is of 2.6 days, while at 26 $^\circ C$ is expected to go to about 4.6 days. The linear fit implies an increase in the doubling time by 50% (or 40%), going from 5 °C to 25 °C., using the estimate from the *extended* dataset (or the *extended* dataset, taking into account of the GDP per capita, at a reference value of 40 thousand dollars).

The main relevance of our study for public health policies, however, is to give motivation to implement safety policies *before* the arrival of the next cold season, in order to avoid a resurgence of the epidemic. For example: keeping social distancing, enforcing the use of masks and providing safe ventilation systems. In buildings where a large

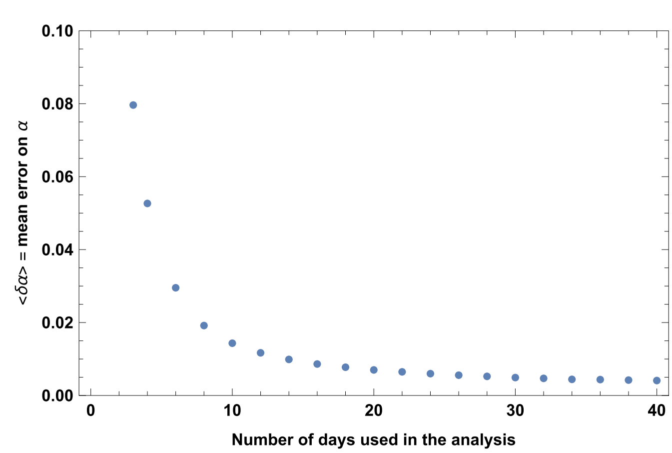


Fig. 8. We show here in the vertical axis the mean value over the full sample of 126 countries of the errors, $\delta\alpha$, on the exponents α, from the exponential fits. On the horizontal axis: the number of days used in the analysis.

[9] Note that a higher serum level of vitamin D has been suggested by many studies and actually found to be correlated with lower virus propagation in Notari and Torrieri (2020), see also McDonnell et al. (2020). However vitamin D is actually quite weakly correlated with Temperature, due to other factors such as diet, as explained in Notari and Torrieri (2020).

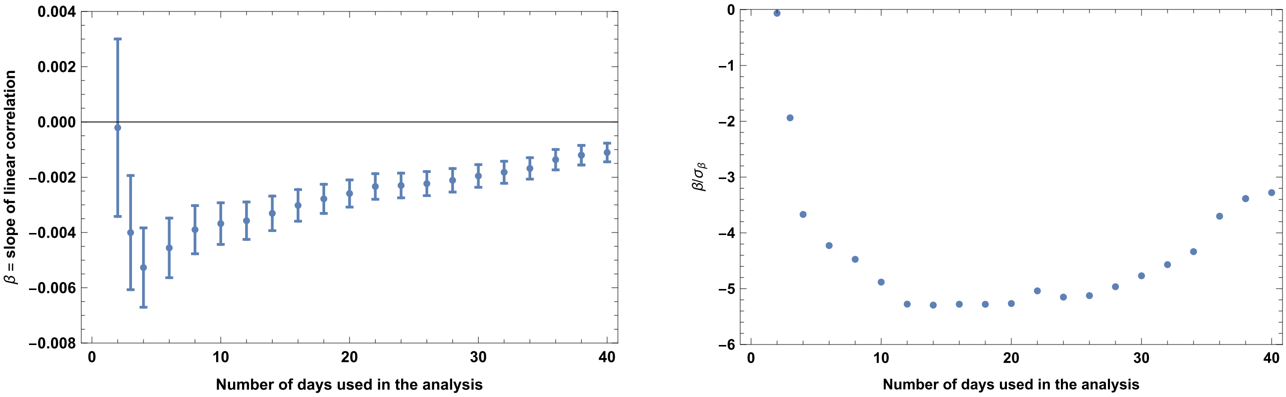


Fig. 9. In the left panel: we show the best estimate for the parameter β of the linear model with the error bar given by its standard error σ_β, as a function of the number of days used in the analysis, after the first day (with a number of cases N_i=30). In the right panel: we show the t-statistic, i.e. β/σ_β, as a function of the number of days used in the analysis.

amount of people are staying together for extended periods of time (such as schools, public offices, workplaces, hospitals) or in transportation systems, a safe ventilation system seems crucial, in order to allow for clean air to enter from outdoors, especially in the cold season. While during mild seasons simply opening windows might be enough to prevent the virus to accumulate in aerosols, this is not possible in the winter in many buildings, unless an efficient and safe ventilation system exists.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

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

References

Adam, Y., et al., 2020. Li. Multivariate analysis of black race and environmental temperature on covid-19 in the US. The American Journal of the Medical Sciences https://doi.org/10.1016/j.amjms.2020.06.015 URL: https://www.amjmedsci.org/article/S0002-9629(20)30257-3/fulltext.

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. doi:https://doi.org/10.1101/2020.03.12.20034728.

Rachel E. Baker, Wenchang Yang, Gabriel A. Vecchi, C. Jessica E. Metcalf, and Bryan T. Grenfell. Susceptible supply limits the role of climate in the early sars-cov-2 pandemic. Science, 369(6501):315–319, 2020. URL: https://science.sciencemag.org/content/369/6501/315, arXiv:https://science.sciencemag.org/content/369/6501/315.full.pdf, doi:https://doi.org/10.1126/science.abc2535.

Bansal S Ryan SJ. Carlson CJ, Gomez ACR. Misconceptions about weather and seasonality must not misguide covid-19 response.

Tomas R. Bolano-Ortiz, Yiniva Camargo-Caicedo, Salvador Enrique Puliafito, Maria Florencia Ruggeri, Sindy Bolano-Diaz, Romina Pascual-Flores, Jorge Saturno, Sergio Ibarra-Espinosa, Olga L. Mayol-Bracero, Elvis Torres-Delgado, and Francisco Cereceda-Balic. Spread of sars-cov-2 through latin america and the caribbean region: a look from its economic conditions, climate and air pollution indicators. Environmental Research, page 109938, 2020. URL: http://www.sciencedirect.com/science/article/pii/S0013935120308331, doi:https://doi.org/10.1016/j.envres.2020.109938.

Qasim Bukhari and Yusuf Jameel. Will coronavirus pandemic diminish by summer? ; available at ssrn: https://ssrn.com/abstract=3556998. 2020. doi:https://doi.org/10.2139/ssrn.3556998.

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. doi:https://doi.org/10.1155/2011/734690.

Jonas Dehning, Johannes Zierenberg, F. Paul Spitzner, Michael Wibral, Joao Pinheiro Neto, Michael Wilczek, and Viola Priesemann. Inferring change points in the spread of covid-19 reveals the effectiveness of interventions. Science, 369(6500), 2020. URL: https://science.sciencemag.org/content/369/6500/eabb9789, arXiv:https://science.sciencemag.org/content/369/6500/eabb9789.full.pdf, doi:https://doi.org/10.1126/science.abb9789.

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. doi:https://doi.org/10.2139/ssrn.3556998.

He et al. Li. Air pollution and temperature are associated with increased covid-19 incidence: A time series study. International Journal of Infectious Diseases, 97:278–282, 2020. URL: https://www.ijidonline.com/article/S1201-9712(20)30383-0/fulltext, doi:https://doi.org/10.1016/j.ijid.2020.05.076.

Huang, Zhongwei, Huang, Jianping, Qianqing, Gu, Pengyue, Du, Liang, Hongbin, Dong, Qing, 2020. Optimal temperature zone for the dispersal of covid-19. Science of The Total Environment 736:139487. https://doi.org/10.1016/j.scitotenv.2020.139487 URL: http://www.sciencedirect.com/science/article/pii/S0048969720330047.

Peter Jüni, Martina Rothenbühler, Pavlos Bobos, Kevin E. Thorpe, Bruno R. da Costa, David N. Fisman, Arthur S. Slutsky, and Dionne Gesink. Impact of climate and public health interventions on the covid-19 pandemic: a prospective cohort study. CMAJ, 192 (21):E566–E573, 2020. URL: https://www.cmaj.ca/content/192/21/E566, arXiv: https://www.cmaj.ca/content/192/21/E566.full.pdf, doi:https://doi.org/10.1503/cmaj.200920.

Liu X et al. Huang J, Zhang L. Global prediction system for covid-19 pandemic.

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. doi:https://doi.org/10.1101/2020.02.12.20022467.

Chandi C. Mandal and M.S. Panwar. Can the summer temperatures reduce covid-19 cases? Public Health, 185:72 – 79, 2020. URL: http://www.sciencedirect.com/science/article/pii/S0033350620302304, doi:https://doi.org/10.1016/j.puhe.2020.05.065.

Mao, Wang, Jiang, Aili, et al., 2020. Temperature Significant Change COVID-19 Transmission in 429 Cities. https://www.medrxiv.org/content/early/2020/02/25/2020.02.22.20025791.

McDonnell S.L.; Baggerly C.A.; French C.B.; Aliano J.L.; Bhattoa H.P. Grant, W.B.; Lahore H. Evidence that vitamin d supplementation could reduce risk of influenza and covid-19 infections and deaths. Nutrients, 12:988, 2020.

Mesay Moges Menebo. Temperature and precipitation associate with covid-19 new daily cases: a correlation study between weather and covid-19 pandemic in oslo, norway. Science of The Total Environment, 737:139659, 2020. URL: http://www.sciencedirect.com/science/article/pii/S004896972033179X, doi:https://doi.org/10.1016/j.scitotenv.2020.139659.

Mugen et al. Ujiie. Effect of temperature on the infectivity of covid-19. International Journal of Infectious Diseases, Volume 95:301–303, 2020. URL: https://www.ijidonline.com/article/S1201-9712(20)30284-8/fulltext, doi:https://doi.org/10.1016/j.ijid.2020.04.068.

Alessio Notari and Giorgio Torrieri. COVID-19 transmission risk factors. medRxiv 2020.05.08.20095083, 2020. doi:https://doi.org/10.1101/2020.05.08.20095083.

Ahmet Ozyigit. Understanding covid-19 transmission: the effect of temperature and health behavior on transmission rates. Infection, Disease & Health, 2020. URL: https://www.idhjournal.com.au/article/S2468-0451(20)30048-1/fulltext, doi:https://doi.org/10.1016/j.idh.2020.07.001.

Shantanu Kumar Pani, Neng-Huei Lin, and Saginela RavindraBabu. Association of covid-19 pandemic with meteorological parameters over singapore. Science of The Total Environment, 740:140112, 2020. URL: http://www.sciencedirect.com/science/article/pii/S0048969720336330, doi:https://doi.org/10.1016/j.scitotenv.2020.140112.

David N. Prata, Waldecy Rodrigues, and Paulo H. Bermejo. Temperature significantly changes covid-19 transmission in (sub)tropical cities of Brazil. Science of The Total Environment, 729:138862, 2020. URL: http://www.sciencedirect.com/science/article/pii/S0048969720323792, doi:https://doi.org/10.1016/j.scitotenv.2020.138862.

Shanna Ratnesar-Shumate, Gregory Williams et al. Simulated sunlight rapidly inactivates SARS-CoV-2 on surfaces. The Journal of Infectious Diseases., jiaa274, 2020. doi:https://doi.org/10.1093/infdis/jiaa274.

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 doi:https://doi.org/10.2139/ssrn.3550308.

Sarmadi, Mohammad, Marufi, Nilufar, Moghaddam, Vahid Kazemi, 2020. Association of covid-19 global distribution and environmental and demographic factors: an updated three-month study. Environmental Research, 188:109748. URL: http://www.sciencedirect.com/science/article/pii/S0013935120306411 https://doi.org/10.1016/j.envres.2020.109748.

Farrukh Shahzad, Umer Shahzad, Zeeshan Fareed, Najaf Iqbal, Shujahat Haider Hashmi, and Fayyaz Ahmad. Asymmetric nexus between temperature and covid-19 in the top ten affected provinces of China: a current application of quantile-on-quantile approach. Science of The Total Environment, 736:139115, 2020. URL: http://www.sciencedirect.com/science/article/pii/S0048969720326322, doi:https://doi.org/10.1016/j.scitotenv.2020.139115.

Peng Shi, Yinqiao Dong, Huanchang Yan, Chenkai Zhao, Xiaoyang Li, Wei Liu, Miao He, Shixing Tang, and Shuhua Xi. Impact of temperature on the dynamics of the covid-19 outbreak in China. Science of The Total Environment, 728:138890, 2020. URL: http://www.sciencedirect.com/science/article/pii/S0048969720324074, doi:https://doi.org/10.1016/j.scitotenv.2020.138890.

Yu Wu, Wenzhan Jing, Jue Liu, Qiuyue Ma, Jie Yuan, Yaping Wang, Min Du, and Min Liu. Effects of temperature and humidity on the daily new cases and new deaths of covid-19 in 166 countries. Science of The Total Environment, 729:139051, 2020. URL: http://www.sciencedirect.com/science/article/pii/S0048969720325687, doi: https://doi.org/10.1016/j.scitotenv.2020.139051.

Jingui Xie and Yongjian Zhu. Association between ambient temperature and covid-19 infection in 122 cities from China. Science of The Total Environment, 724:138201, 2020. URL: http://www.sciencedirect.com/science/article/pii/S0048969720317149, doi: https://doi.org/10.1016/j.scitotenv.2020.138201.