The Influence of Weather Conditions on 2019-nCovid using Geospatial Data

Francois Cohen[i], Sihan Li[ii], Yangsiyu Lu[i] and Moritz Schwarz[i,iii]

This version: 1st April 2020; first version: 1st April 2020; data update: March 29th, 2020.

Supplementary material

1. Statistical model

We use the following econometric model to estimate the impact of the weather on infections:

(1) $\ln(C_{i,t}) - \ln(C_{i,t-1}) = \sum_{x=0}^{X} a_x . W_{i,t-x} + \mu_{c,t} + n_i + \epsilon_{i,t}$

In Eq. (1), $\ln(C_{i,t})$ is the logarithm of the total number of confirmed cases of 2019-nCov observed in area i on day t. The dependent variable is therefore the first difference of this logarithm. This transformation allows us to scale any change in infections in relative terms based on the level of infections the day before. This is to account for the fact that infections can only be proportional to the number of people already infected in an area.

$W_{i,t-x}$ is a matrix of weather-related variables that includes information on the weather at time $t - x$. We include the lagged values of these weather variables (until $t - X$) to capture the effect of the weather of the previous days on infections. $W_{i,t-x}$ is modulable. We run specifications with temperatures, humidity and precipitation. We often separate average temperatures into minimum and maximum temperatures. We also use different values for the total number of lags (X). Our main models use 15 lags ($X = 15$) and therefore covers 16 days. This should cover the maximum time reported for the incubation of the disease (14 days) and its detection through testing. We report alternative models with less lags, and more lags, in **supplementary material 3**.

$\mu_{c,t}$ are country by day fixed effects (e.g. the UK on March 25th, 2020). They therefore control for national factors which may vary from day to day and influence the spread of the disease. n_i is an area-specific (e.g. regions or cities) fixed effect and $\varepsilon_{i,t}$ is the error term. The parameters a_k are the vectors of interest to be estimated. With this specification, the effect of temperature on the spread of 2019-nCov is identified from sub-regional deviations in confirmed cases within a country and within a day ($\mu_{c,t}$), accounting for the average differences in the number of cases between areas (controlled by n_i). The model is estimated using the estimator developed by Correia (2018)[1]. We cluster standard errors at the country level.

We tried another specification before opting for Eq. (1):

(2) $c_{i,t} = \exp\left(\sum_{x=0}^{X} a_k . W_{i,t-x} + \mu_{c,t} + n_i\right) + \epsilon_{i,t}$

In Eq. (2), $c_{i,t}$ is the number of confirmed cases of 2019-nCov observed in area i, on day t. All other variables remain unchanged. This specification is a Poisson model and often used to look at count data. The main problem with this specification is that it did not converge when we included the many lags.

---

[i] Smith School of Enterprise and the Environment, University of Oxford; and Institute for New Economic Thinking at the Oxford Martin School, University of Oxford.
[ii] Environmental Change Institute, University of Oxford.
[iii] Climate Econometrics, Nuffield College, Oxford

We therefore ended up choosing a specification that is very close to a count model in principle but offers the advantage of being linear.

2. Robustness checks

Choice of fixed effects. The specifications below illustrate the importance of using fixed effects to isolate the impact of temperatures on infections. Models (1) to (3) strongly underestimate 2019-nCov response. Model (4) provides similar results to the model of Table 1, column (1), here reproduced in column (5). This unsurprisingly suggests that models that do not include day fixed effects or use country fixed effects instead of area fixed effects are insufficient to estimate the response of 2019-nCov infections to the weather.

Appendix Table A1: Robustness checks on the choice of the fixed effects

Column	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Av. Temperature (°C)	-0.0007	-0.0006	-0.0001	-0.0042***	-0.0031***
	(0.0005)	(0.0005)	(0.0002)	(0.0015)	(0.0006)
Precipitations (mm)	0.0038*	0.0043**	0.0009	0.0014	0.0002
	(0.0019)	(0.0019)	(0.0011)	(0.0022)	(0.0014)
Relative humidity (%)	0.0002	-0.0001	-0.0001	-0.0007	-0.0005***
	(0.0007)	(0.0006)	(0.0002)	(0.0006)	(0.0001)
Fixed effects (Y/N):					
Day	N	Y	Y	Y	Y
Country	N	N	Y	Y	Y
Area	N	N	N	Y	Y
Day by country	N	N	N	N	Y

Notes: Alternatives to the model displayed in Table 1, column 1, which a different choice of fixed effects. Standard errors are in brackets and clustered at country level. *, **, and *** are for statistical significance at 10%, 5% and 1% respectively.

Model dynamics. We change the number of lags below. The effect of temperature tends to grow with the number of lags included in the model. This is because fewer infections in the past leads to even fewer infections today. Good weather will have an accumulative effect on reducing the number of coronavirus cases.

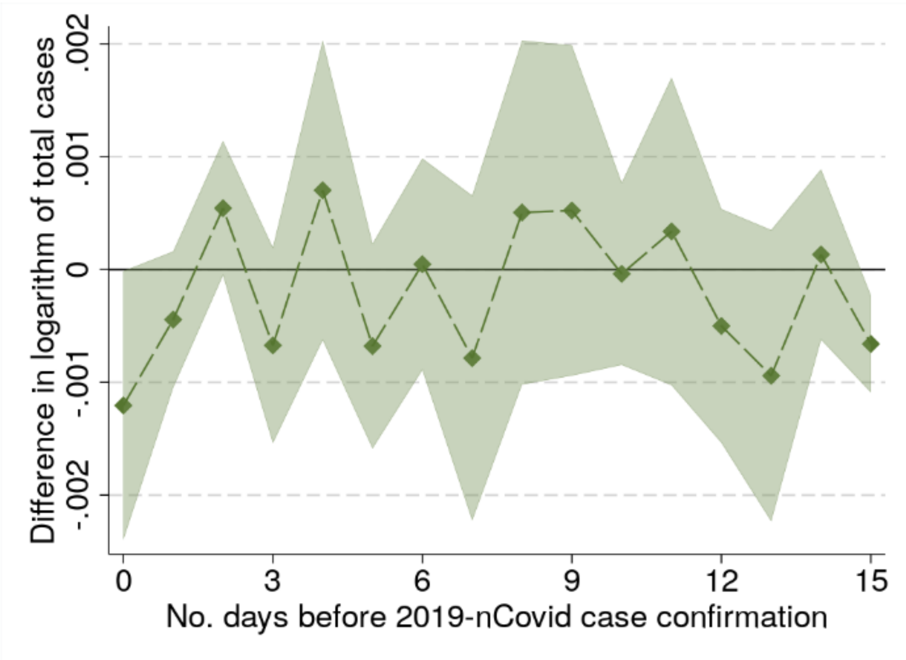
Appendix Table A2: Alternative number of day lags

	Average Temperature (°C)	
Specification	Coefficient	Standard error
20 daily lags	-0.00416***	(0.00048)
19 daily lags	-0.00346***	(0.00037)
18 daily lags	-0.00334***	(0.00041)
17 daily lags	-0.00332***	(0.00039)
16 daily lags	-0.00314***	(0.00039)
14 daily lags	-0.00313***	(0.00058)
13 daily lags	-0.00291***	(0.00053)
12 daily lags	-0.00261***	(0.00074)
11 daily lags	-0.00193**	(0.00074)
10 daily lags	-0.00123**	(0.00050)
9 daily lags	-0.00133***	(0.00042)
8 daily lags	-0.00118**	(0.00045)
7 daily lags	-0.00168***	(0.00044)
6 daily lags	-0.00209**	(0.00079)
5 daily lags	-0.00168***	(0.00026)
4 daily lags	-0.00185***	(0.00018)
3 daily lags	-0.00151***	(0.00025)
2 daily lags	-0.00160***	(0.00052)
1 daily lag	-0.00143***	(0.00053)
No lag	-0.00148*	(0.00074)

Notes: All the specifications are based on Table 1, column 1 by using different number of daily lags. The coefficients are for the lagged temperatures combined. Standard errors are in brackets and clustered at country level. *, **, and *** are for statistical significance at 10%, 5% and 1% respectively.

The graphs and figures included in the main text are for the cumulative effect of all the lags included in the model (the sum $\sum_{x=0}^{X} a_x$.). The graph below furthermore reports the individual effects (each single a_x) for temperature and the model of **Table 1, column (1)**. Coefficients are naturally estimated noisily. However, we find an impact of on-the-day temperature on the number of confirmed cases, suggesting that colder weather may not only play a role in contagion, but also in how the immune system is fighting the disease.

Appendix Figure A1: Values of the individual coefficients (a_x) for daily average temperatures in model (1)



Newly infected countries. Below we check if the spread of the virus correlates with the weather at the start of the contagion, when a country records less than a given number of total cases (as estimated in the dataset). We find an association in the early stages of the disease, but it is inefficiently estimated.

Appendix Table A3: Restricting the sample to below a certain number of national cases

Column	(1)	(2)	(3)	(4)	(5)
Only in sample if total cases in country are:	<500	<1000	<2000	<5000	All data
Av. Temperature (°C)	-0.0094*	-0.0051	-0.0069	-0.0021	-0.0031***
	(0.0053)	(0.0047)	(0.0042)	(0.0042)	(0.0006)
Precipitations (mm)	-0.0031	-0.0032	-0.0018	-0.0024	0.0002
	(0.0021)	(0.0022)	(0.0022)	(0.0020)	(0.0014)
Relative humidity (%)	-0.0011	-0.0014	-0.0014	-0.0001	-0.0005***
	(0.0016)	(0.0013)	(0.0013)	(0.0014)	(0.0001)
Observations	10,995	15,269	17,109	21,283	55,712

Standard errors are in brackets and clustered at country level. *, **, and *** are for statistical significance at 10%, 5% and 1% respectively.

3. Summary statistics of meteorological data

Appendix Table A4: Summary statistics of the meteorological data after it is matched to the 2019-nCov data

Variable	Mean	Std. Dev.	Min	Max
Av. temperature (°C)	8.74	8.76	-30.81	38.08
Min. temperature (°C)	4.10	9.43	-36.25	32.62
Max. temperature (°C)	13.39	8.53	-29.00	43.55
Total Precipitation (mm)	2.20	5.05	0.00	128
Relative humidity (%)	67.49	17.79	5.74	99.97
Temperature bins				
Average temperature:				
<0°C	0.13	0.34	0	1
0-10°C	0.47	0.5	0	1
10-20°C	0.28	0.45	0	1
20-30°C	0.12	0.32	0	1
>30°C	<0.01	0.04	0	1
Min. temperature:				
<-10 °C	0.06	0.24	0	1
-10-0°C	0.26	0.44	0	1
0-10°C	0.45	0.5	0	1
10-20°C	0.16	0.37	0	1
>20°C	0.07	0.25	0	1
Max. temperature:				
<10°C	0.35	0.48	0	1
10-20°C	0.44	0.5	0	1
20-30°C	0.19	0.39	0	1
30-40°C	0.02	0.15	0	1
>40°C	<0.01	0.01	0	1

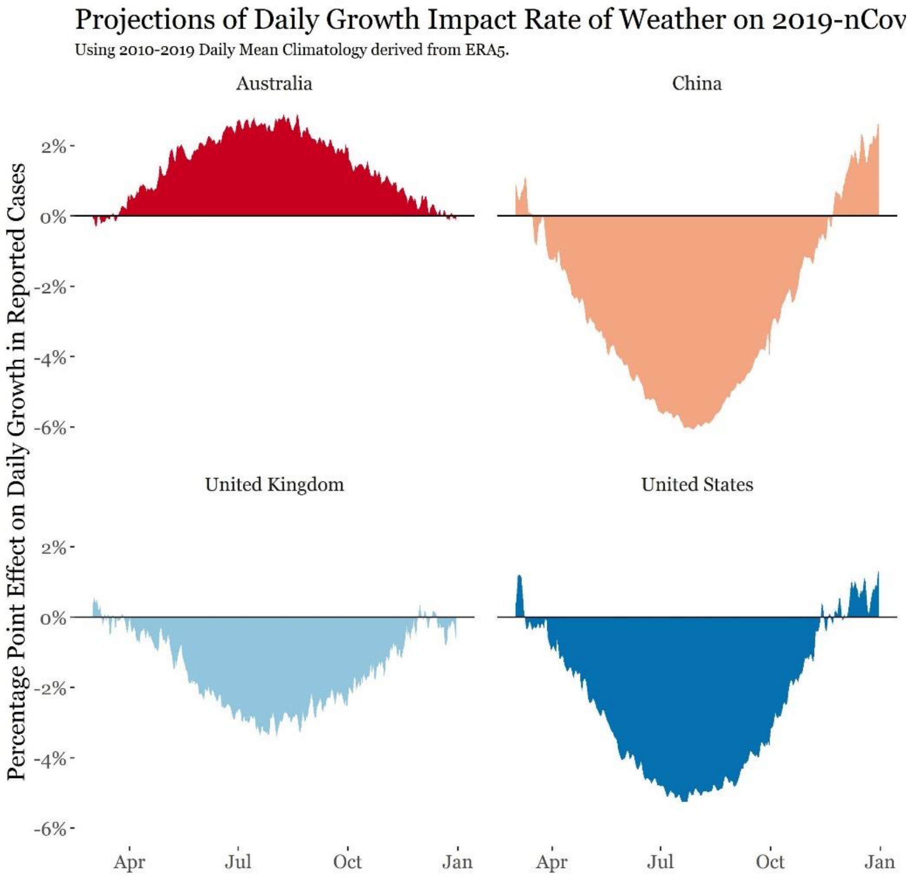
4. Projection Details

The projections have been produced in four steps. First, we use the 10-year average (2010-2019) of daily temperature and relative humidity from ERA5 for March to December and aggregate this data to the country-level using 2020 population weighting.[2] We construct a daily anomaly dataset of the weather conditions for March 1st –December 31st with respect to the monthly mean values for March, as most confirmed cased considered in this study have been observed in March this year.

This gives us a very rough estimate of the difference in the expected weather conditions by country with respect to March until the end of 2020.

Third, we use the estimates of Table 1, column 1 to predict changes in the average growth rate of 2019-nCov confirmed cases. These impacts on the daily growth rate are provided in the Figure A2 below.

Appendix Figure A2: Changes in the daily growth rate of confirmed cases as a function of the expected weather from 1 March 2020



Fourth, we insert these estimates of the weather impacts on the daily growth rate of infection into a simple susceptible-infectious-recovered (SIR) compartment model,[3,4,5] using the parameters provided in Walker et al. (2020) for 2019-nCov. This allows us to investigate the evolution of the disease in three social distancing scenarios (no distancing, distancing for the elderly, and distancing for the whole population), as a function of the other parameters in the model, especially the value of the R_0, and the weather.

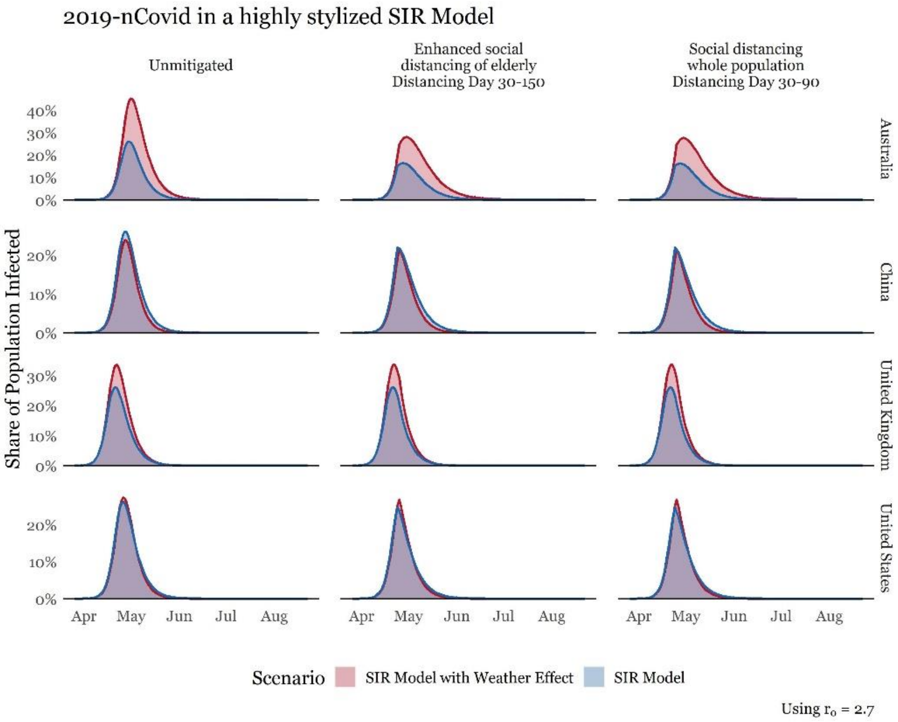
We also rely on Walker et al. (2020)**Error! Bookmark not defined.** for their parameters of R_0, which are 2.4, 2.7, 3 and 3.3. We provide the results for 2.4 in the core of the text, and the projections for other values hereafter.

The two social distancing scenarios are the ones given by Walker et al. (2020). We use the social contact rates calculated by these authors. Social distancing measures are assumed to come into effect 30 days after the start of the model. To simulate a relaxation of social distancing measures, the scenario that assumes social distancing to the entire population keeps their measures only in place for 60 days and relaxes them again after that. Walker et al. (2020) use scenario-specific contact reduction rates of up to 48%.

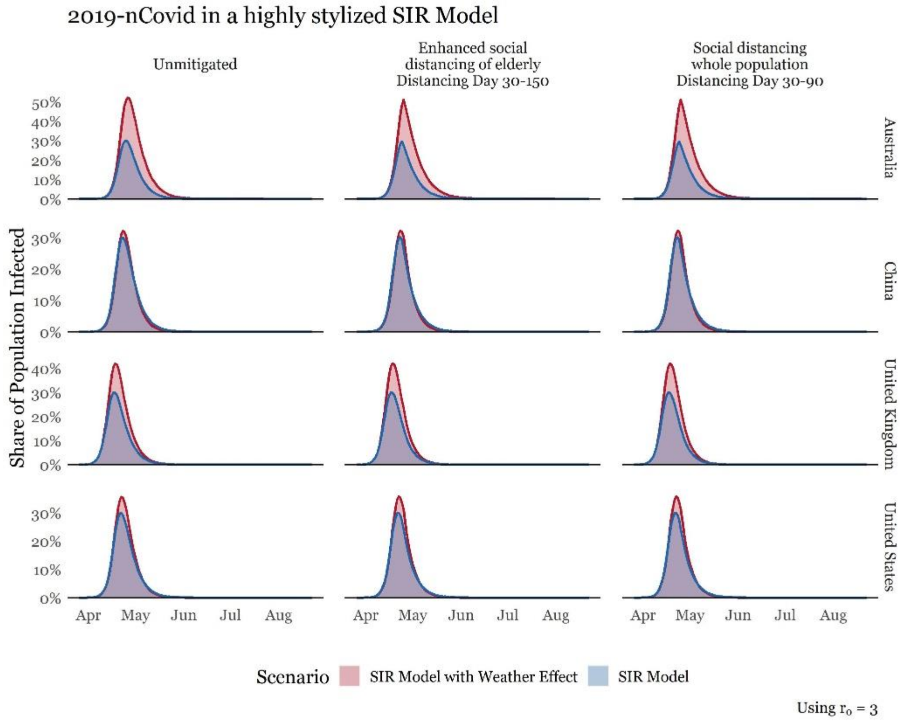
The initial number of cases for each country is the latest recorded date in the Xu et al. (2020) database at the time of writing. Each model is run for 150 days.

The growth projections described above were added to the endogenous daily SIR growth rates after each model was run to produce the weather effect. This constitutes only a first attempt, as more sophisticated insights could be developed by endogenizing the weather effect into the SIR model.

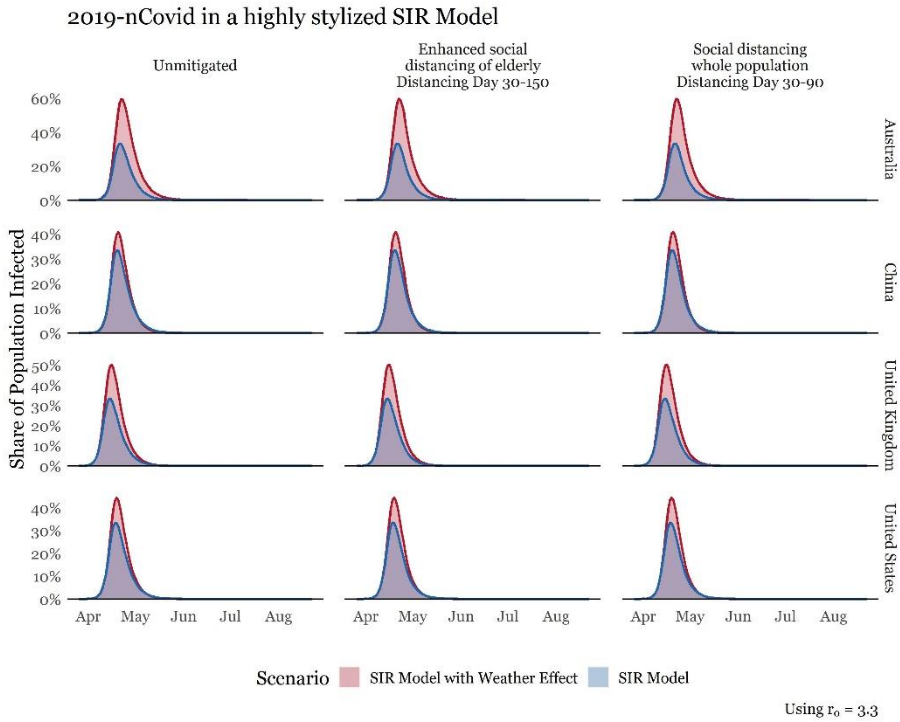
Appendix Figure A3: Output of SIR model with R0 = 2.7



Appendix Figure A4: Output of SIR model with R0 = 3



Appendix Figure A5: Output of SIR model with R0 = 3.3



ADDITIONAL REFERENCES

---

[1] Correia, S. (2018). REGHDFE: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects.

[2] Center for International Earth Science Information Network - CIESIN - Columbia University. 2016. Gridded Population of the World, Version 4 (GPWv4.11): Population Count. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H4X63JVC . Accessed [1st April, 2020].

[3] Schneider, T. (2020). Flatten the Curve. Code available at: https://github.com/tinu-schneider/Flatten_the_Curve [Accessed 30th March 2020].

[4] Höhle, M. (2020). Flatten the COVID-19 curve. Blog post available at: https://staff.math.su.se/hoehle/blog/2020/03/16/flatteningthecurve.html [Accessed 30th March 2020].

[5] Kermack, W. O., and A. G. McKendrick. (1927). A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society, Series A* 115: 700–721.