Ultraviolet radiation decreases COVID-19 growth rates: Global causal estimates and seasonal implications

Tamma Carleton, Jules Cornetet, Peter Huybers, Kyle C. Meng, and Jonathan Proctor*

This version: May 15, 2020
First version: April 20, 2020

*Carleton: University of Chicago, Dept. of Economics and Energy Policy Institute (email: tcarleton@uchicago.edu). Cornetet: École Normale Supérieure Paris-Saclay, Département de Sciences Sociales (email: jules.cornetet@ens-paris-saclay.fr). Huybers: Department of Earth and Planetary Sciences, Harvard University (email: phuybers@fas.harvard.edu. Meng: Bren School of Environmental Science and Management, Dept. of Economics, and Environmental Markets Solutions Lab (emLab), UC Santa Barbara and National Bureau of Economic Research (email: kmeng@bren.ucsb.edu). Proctor: Center for the Environment and Data Science Initiative, Harvard University (email: jproctor1@fas.harvard.edu). Equal contributions. We thank our families for their patience with this project during a trying period. We received helpful comments and feedback from Max Auffhammer, Alan Barreca, Olivier Deschenes, Jonathan Dingel, Nick Hagerty, Solomon Hsiang, Marc Lipsitch, Gordon McCord, Antony Millner, Fran Moore, Ishan Nath, James Rising, Ashwin Rode, Mauricio Santillana, and Geoffrey Schiebinger. Borui Sun and Shopnavo Biswas provided excellent research assistance. The authors declare no conflicts of interest.

Abstract

Nearly every country is now combating the 2019 novel coronavirus (COVID-19). A central concern is whether COVID-19 transmission exhibits seasonality. If so, changing environmental conditions in coming months may shift COVID-19 infection patterns and policy responses around the world. We estimate a relationship between growth of confirmed cases of COVID-19 and local environmental conditions by combining the most spatially-disaggregated global dataset of daily cases assembled to date, consisting of 3,235 administrative units across 173 countries, with a statistical model isolating random variation in daily weather conditions. Our analysis indicates a strong effect of ultraviolet radiation (UV) on daily COVID-19 growth rates: a 1 kJ m^{-2} increase in hourly UV decreases the growth rate of confirmed COVID-19 cases by .09 percentage points ($\pm$.04, p =.01), with a delayed effect that manifests over two weeks. We find weak or inconsistent lagged effects of local temperature, specific humidity, and precipitation. To illustrate the seasonal implications of our findings with respect to UV, we show that estimated UV effects imply decreases in COVID-19 growth rates of 1.18 percentage points ($\pm$0.47) in the extra-tropical Northern Hemisphere and increases of 2.1 percentage points ($\pm$0.83) in the extra-tropical Southern Hemisphere between April and July 2020. Seasonality in UV dictates the inverse pattern for January 2021, with COVID-19 growth rates rising by 5.5 percentage points ($\pm$2.18) in the extra-tropical Northern Hemisphere and falling by 4.82 percentage points ($\pm$1.91) in the extra-tropical Southern Hemisphere, relative to April 2020. These effects are substantial when compared to the average in-sample COVID-19 growth rate of 13.21 percent. The total seasonal effect of all climate variables investigated is indeterminate in sign due to uncertainty in the effects of temperature and specific humidity. Although many factors will influence future COVID transmission, our findings suggest a need for adjustment of COVID-19 containment policies for the seasonality of UV.

Introduction

In late 2019, a novel virus strain from the family *Coronaviridae*, referred to as SARS-CoV-2, began spreading throughout China [1]. Central among SARS-CoV-2 concerns are its high transmissivity, potentially severe flu-like symptoms, and high case fatality rates [2]. In the ensuing months, the virus has spread globally, prompting the World Health Organization to declare a pandemic on March 11, 2020. At the time of this writing, cases of COVID-19, the disease caused by SARS-CoV-2, have been detected in almost every country (Fig. 2A), with the number of confirmed global cases in the millions.

Much remains unknown about COVID-19. An important question concerns how environmental conditions modify COVID-19 transmission. In particular, sensitivity to environmental conditions that vary seasonally may allow prediction of transmission characteristics around the globe over the coming months and have implications for seasonal re-emergence of infections [3]. Previous studies show that prevalence of H3N2, 2009 H1N1, and other strains of influenza is sensitive to local temperature and specific humidity [4, 5, 6, 7, 8]. Given that ultraviolet radiation (UV) can damage nucleic acids generally, and inactivate related strains of coronavirus and influenza [9, 10, 11], it has been hypothesized that a similar sensitivity exists for SARS-CoV-2. Indeed, SARS-CoV-2 sensitivity to UV exposure is suggested by the widespread use of UV for disinfecting medical equipment [12, 13].

Existing efforts to understand the potential seasonality of COVID-19 face several major hurdles. First, a number of different environmental variables have been postulated to affect transmission including temperature, precipitation, specific or relative humidity, and UV [14, 15, 16, 17, 18]. These atmospheric variables are dynamically linked (Fig. S4). For instance, solar radiation is correlated with relative humidity and precipitation through cloud formation and convection. This, in turn, is correlated with cooling, increased specific humidity, and decreased UV radiation at the surface. Co-variation among these environmental variables makes it difficult to discern individual contributions.

Second, most studies to date compare environmental conditions and COVID-19 cases or growth rates across locations. Such cross-sectional analyses conflate location-specific characteristics – such as testing capacity, population density, and health services – with environmental conditions. Because environmental conditions are often correlated with unobservable confounding factors, such cross-sectional comparisons may not have a causal interpretation [19, 20, 21]. For example, countries that are cooler on average also tend to have higher income per capita [22], with the latter feature associated with more widespread testing, hospital access, and potential for identifying COVID-19 cases. Reacting in part to concerns over confounding factors in prior studies, a recent review by the National Academies of Sciences, Engineering and Medicine observed that temperature and humidity effects on COVID-19 remain inconclusive [23].

Finally, any modification of transmission will appear only subsequently in observations of confirmed COVID-19 cases. The incubation period of COVID-19 is estimated to average 6 days and range from 2 to 14 days [24, 25, 26, 27, 28], with additional delays for testing and reporting. Reduced growth rates of COVID-19 cases occurred 8 to 12 days after nation-wide restriction in travel in China [29]. Most studies have assumed either an instantaneous response, or a response to averaged environmental conditions over an immediately preceding interval. Estimation of an accurate and causal response requires flexibly accounting for the lag between transmission and confirmation of COVID-19 cases.

This study employs statistical techniques that isolate random local fluctuations in environmental conditions to estimate plausibly causal global relationships between the daily growth rate of confirmed COVID-19 cases, hereafter the "COVID-19 growth rate," and preceding local temperature, specific humidity, precipitation, and UV [21]. Detecting the effects of local environmental conditions requires a harmonized global

datatset of geolocated COVID-19 confirmed cases, hereafter "cases", at relatively high spatial resolution. We assemble what is, to our knowledge, the most spatially disaggregated global dataset of COVID-19 cases to date by combining subnational and national daily COVID-19 case data (Supplementary Information Section B, Tables S2, S3). Our data consist of a total of 1,153,726 COVID-19 cases from 3,235 geospatial units covering 173 countries and five continents (Fig. 1A). We focus on the growth rate of confirmed cases rather than that of infected individuals because data on recoveries and deaths is not consistently available globally. These data span January 01, 2020 to April 10, 2020, though coverage across locations is uneven during this period. Before March 2020, China, South Korea, and Japan are heavily represented. Since March 2020, coverage is nearly global (Fig. 1B). We cross-reference data across multiple sources when available (Supplementary Information Section B and Fig. S1).

We construct our outcome variable using the first difference in the natural logarithm of daily confirmed cases. We report growth rates as percentage changes, which is a linear approximation of the daily change in the logarithm of cases. The mean daily growth rate in our sample period is 13.21% with a standard deviation of 24.36%. Generally, the COVID-19 growth rate is larger at the beginning of the epidemic and decreases over time: the sample average growth rate falls from 15.66% during the first month of a location's outbreak to 2.31% afterwards (Fig S2). Decreases in growth rate are consistent with strengthened containment efforts over time, though may also reflect other factors. We merge these daily COVID-19 growth rate records with daily temperature, specific humidity, precipitation, and UV for each national and subnational geospatial unit. These environmental variables were generated by a state-of-the-art climate reanalysis model in near real-time at the a 0.25° latitude by 0.25° longitude resolution and then aggregated to each geospatial unit using population weights [30, 31].

The spatial and temporal coverage of this global longitudinal dataset enables us to estimate empirical relationships between local environmental conditions and COVID-19 infection in a manner that addresses shortcomings of previous work. First, by jointly estimating the effects of UV, temperature, specific humidity and precipitation in a single regression model, we recover the effect of each variable, accounting for variation in the others (Supplementary Information Section A.2). Second, we exploit the longitudinal structure of the dataset to examine the time period between initial COVID-19 transmission and COVID-19 case confirmation through the use of a temporal distributed lag model. COVID-19 growth rates are specified as a linear function of UV, temperature, specific humidity, and precipitation over the preceding 17 days, allowing the effect of each environmental variable to differ in each of six 3-day lag intervals (Supplementary Information Section A.2). We summarize this dynamic response to each environmental variable by summing lagged coefficients to recover cumulative effects. As a check of our ability to recover the influence of environmental shocks on transmission and their eventual effect on case growth rates, we estimate a distributed lag regression model using data simulated by a standard stochastic Susceptible-Exposed-Infected-Recovered (SEIR) model [32] (Supplementary Information Section A.1 and Fig. S11). We note that our statistical estimate includes both biological and human-behavioral channels through which environmental conditions may alter COVID-19 transmission. Public health responses to COVID-19 require understanding the combined effects of both channels.

Finally, the spatial and temporal coverage of our data enables us to follow recent advances in the climate econometrics literature [21] in employing a suite of semi-parametric controls designed to isolate the effects of local temperature, specific humidity, precipitation, and UV from potentially confounding factors that may also influence the growth of COVID-19 cases. These controls account for time-invariant differences in population characteristics across countries, such as differential population densities and health care systems;

temporal shocks that influence the pattern of global COVID-19 events, such as the WHO pandemic designation; and seasonal time trends that may vary within each national and subnational unit (Fig. S3). Together, these controls isolate idiosyncratic daily variations in environmental conditions experienced by the same population. Results are consistent across a range of alternative statistical models that vary the stringency of these controls (Supplementary Information Section A.2).

Results

We find robust statistical evidence that increased UV lowers the subsequent COVID-19 growth rate. We estimate that 1 kJ/(m^2 hour) increase in local UV reduces the COVID-19 growth rate by .09 percentage points over the course of 17 days, with a 95% confidence interval of [-.16, -.02] (Fig. 2A and col. 2 of Table S1). The effects associated with UV are consistently negative across lags, and peak in magnitude at a delay of 9-11 days. This one- to two-week period between UV exposure and changes in the COVID-19 growth rate is consistent with the reported time frame between exposure to the virus and its detection [26, 33, 34]. In contrast, the effects of higher temperatures and higher levels of specific humidity are of less consistent sign, with cumulative effects over the 17-day interval being statistically insignificant and of opposite sign than that of the lag with the greatest magnitude (Fig. 2B, second and third columns). The estimated UV effects imply that a sample standard deviation increase in UV (14.42 kJ/(m^2 hour), accounting for semi-parametric controls), reduces growth by 1.31 percentage points, relative to a sample average growth rate of 13.21%, with a 95% confidence interval of [-2.31, -.3] percentage points. This amounts to a rise in the doubling time of COVID-19 cases from 5.25 days – at the average growth rate – to 5.82 days, with a 95% confidence interval of [5.33, 6.32] days. A standard deviation increase in temperature (3.55 °C) and specific humidity (0.16 %) leads to small and uncertain estimated effects on the growth rate of 0.61 and 0.35 percentage points, respectively.

The effect of UV on the COVID-19 growth rate (Fig. 2A) is robust to a range of alternative statistical models, including controls for days since initial outbreak of the virus in each location (col. 1, Table S1); linear, country-specific time trends (col. 2, Table S1); controls for future weather as a test of reverse causality (col. 4, Table S1); controls for temporally and spatially-varying policies, such as work-from-home policies, school closures and event cancellations (col. 5, Table S1 and col. 2, Fig. S6); controls for the extent of COVID-19 testing availability (col. 6, Table S1); semi-parametric controls that allow for within-country differential seasonality, in addition to country-wide seasonal patterns (col. 7, Table S1 and col. 3, Fig. S6); removal of country-level data and addition of semi-parametric controls that allow for country-by-day specific shocks (col. 8, Table S1 and col. 4, Fig. S6); estimation of a Poisson model (Eq. S2 in Supplementary Information Section A.2) in place of ordinary least squares (col. 9, Table S1 and col. 5, Fig. S6); replacing specific humidity with relative humidity (Fig. S7); and changing the number of lags included in the estimating equation (Fig. S8). Findings associated with temperature and humidity are similarly unresponsive to changes in specification.

The dynamic effects of weather conditions on COVID-19 growth rates shown in Fig. 2A reflect average treatment effects over all geospatial units and over the course of the observed pandemic. We find suggestive evidence, however, that social distancing policies (school closures, mandatory work from home orders, and large event cancellation regulations) weaken the link between COVID-19 and weather conditions. Using a binary policy variable indicating whether an administrative unit has any one of a set of social distancing measures in place (Supplementary Information Section B.3), we find that the cumulative effect of 1 kJ/(m^2 hour) increase in UV falls from -.09 ($\pm$.04, p=.02) before policies are introduced to -.04 ($\pm$.06, p=.52) after

policies are put in place (Fig. 2B). This difference, however, is not statistically significant (p=.47). While a similar pattern is observed for the individual lags of temperature and specific humidity that have the largest magnitude (Fig. S9), cumulative effects over the 17-day period are statistically indistinguishable from zero both with and without public health policies in place (Fig. 2B). Similarly, we find suggestive evidence that the cumulative effect of UV on the growth rate of COVID-19 cases is stronger during the first month of a population's outbreak (-.13±.04, p=.003), than after the first month of outbreak (-.08±.04, p=.02). Again, a similar pattern is observed for the individual lags of temperature and specific humidity that have the largest magnitude, but cumulative effects are statistically indistinguishable from zero across the entire sample period (Figs. 2B and S9).

The estimated effect of UV on the COVID-19 growth rate has seasonal implications (Fig. 3). To illustrate these, we use the cumulative effect of UV recovered in Fig. 2B, along with seasonal climatology of UV, to simulate differential growth rates under alternative climatologies (Supplementary Information Section A.3). Differences in simulated COVID-19 growth rate between average conditions in July and April generally correspond with changes in incoming solar radiation (Fig. S5). Specifically, we estimate that in July, relative to April, extra-tropical Northern Hemisphere COVID-19 growth rates decrease by 1.18 (±0.47) percentage points, whereas they increase by 2.1 (±0.83) in the extra-tropical Southern Hemisphere (Fig. 3A-B). This seasonal change amounts to an increase in the doubling time from an average of 5.25 to 5.77 days in the extra-tropical Northern Hemisphere and a corresponding decrease to 4.52 days in the extra-tropical Southern Hemisphere. Conversely, comparing January to April in our simulation indicates that COVID-19 growth rates are 5.5 (±2.18) percentage points higher in the extra-tropical Northern Hemisphere, and 4.82 (±1.91) percentage points lower in the extra-tropical Southern Hemisphere (Fig. 3C-D); this corresponds to lowering the average doubling time to 3.69 days in northern latitudes and raising it to 8.31 days in southern latitudes. Daily simulations for every country over the next 11 months further reveals the temporal pattern of UV-driven seasonality in COVID-19 growth rates (Figs. 3E-F, S10). As a whole, the tropics have a moderate seasonal change in the daily growth rate of -0.56 (±0.22) percentage points between April and July and -1.12 (±0.44) percentage points between April and January (Fig. 3B,D), though migration of the inter-tropical convergence zone and associated changes in cloudiness has regional implications. A notable exception to these hemispheric-wide trends is that the onset of the South Asian monsoon causes decreased surface UV regionally in July, thus raising COVID-19 risks in coming months.

The fact that UV closely follows the seasonal cycle of solar radiation makes it distinct from from many weather variables whose seasonality lags solar radiation by as much as several months [35]. A similar illustrative exercise using temperature or specific humidity is indeterminate in sign because the cumulative effect of each variable is uncertain (Fig. 3B,D). The uncertain contribution of these variables renders the total effect of seasonality across all three variables uncertain, with all 95% confidence intervals for the cumulative effect containing zero (black bars, Fig. 3B,D).

We emphasize that simulations of UV-driven seasonality in COVID-19 are merely illustrations of one component of seasonality in the epidemic. Other climate variables may also influence COVID-19 infection patterns in the coming months, including seasonal changes in temperature and specific humidity, whose effects on COVID-19 transmission rates remain uncertain. Moreover, climatological conditions themselves are only one of many factors that will determine global COVID-19 prevalence in the coming months [36].

Discussion

Using a global, harmonized dataset of daily COVID-19 cases assembled at what we believe is the finest spatial resolution to date, we find that the daily growth rate of confirmed COVID-19 cases responds negatively to increased UV. Importantly, variations in the COVID-19 growth rate lag variations in UV by as much as two weeks, consistent with the time required for incubation, testing, and reporting. The UV response is robust to a range of model specifications and controls. Other environmental factors that are jointly examined are not found to meaningfully influence COVID-19 cases. The distributed lag panel regression framework that we employ here will be useful for future analyses of the influence of other environmental conditions, particularly as COVID-19 data availability improves.

Although our statistical approach does not isolate effects of UV on transmission alone, our findings are consistent with the hypothesis that UV radiation alters COVID-19 transmission rates by more-rapidly deactivating the SARS-CoV-2 virus residing on surfaces or in aerosol form, as suggested by the general UV-sensitivity of nucleic acids [37]. Simulations from an SEIR model show that if UV influences the growth rate of cases through only the transmission rate, our statistical estimate corresponds to an underestimate of the effect of UV on transmission (Supplementary Information Section A.1). However, as it is possible that UV also influences the incubation period of SARS-CoV-2, testing rates, or other properties of the disease; future research should provide a more direct mapping between statistical estimates and key parameters found in process-based models such as the SEIR model.

Our study has a number of important limitations. First, as is true in any empirical study of disease, we can only observe cases that are confirmed, and it is very likely that confirmed cases of COVID-19 fall far below the actual number of infections [38]. Although persistent heterogeneity in testing rates across administrative units would not influence growth rate estimates, administrative units have also changed testing procedures over time [39]. We address this concern statistically by accounting for location-specific trends in confirmed COVID-19 cases and by controlling for the availability of COVID-19 testing at the country level (Supplementary Information Section B and Table S1), though reporting issues could remain.

Second, it is possible that the behavioral response to random day-to-day fluctuations in UV differs from the behavioral response to expected seasonal changes. That is, an anomalously sunny day in March may elicit different human behavior than a day in July with the same UV exposure. Similarly, the ability of an increase in UV to further inhibit transmission of SARS-CoV-2 may decrease at higher baseline UV levels. Only after the COVID-19 pandemic progresses further and data across multiple seasons becomes available will it be possible to empirically address this concern. While behavioral factors may render short- and longer-run responses to UV different, we find using stochastic simulations of the SEIR model that this is not the case if UV influences transmission alone. Using simulated data, we show that the estimated cumulative effect of weather on the daily growth rate of confirmed cases is only weakly sensitive to the frequency structure of simulated weather-induced shocks to transmission (Supplementary Information Section A.1 and Fig. S12).

Third, some studies suggest a relationship between air pollution and COVID-19 transmission and mortality [40, 41, 42, 43]. Although it is theoretically possible that the negative effect of UV that we recover is partially explained by air pollution attenuating UV and increasing COVID transmission, this is unlikely to be the case, given that day-to-day variation in UV is driven primarily by changes in cloud cover, with minor contributions coming from variations in ozone, aerosols, and water vapor [44].

Finally, although we know of no publicly available laboratory studies of the UV-COVID-19 transmission relationship to date, we view our approach as complementary to such future results. Although laboratory studies isolate the biology of virus transmission, our statistical approach using observed COVID-19 cases

captures those channels *as well as* any behavioral adjustments individuals make in response to short-term UV fluctuations, such as decisions to go outside, to exercise, to attend social gatherings, and many other possible activities and health investments [45]. As public health officials grapple with the costs and benefits of a range of possible responses to the current pandemic, quantifying the influence of both channels is essential to building appropriate policies.

Figures

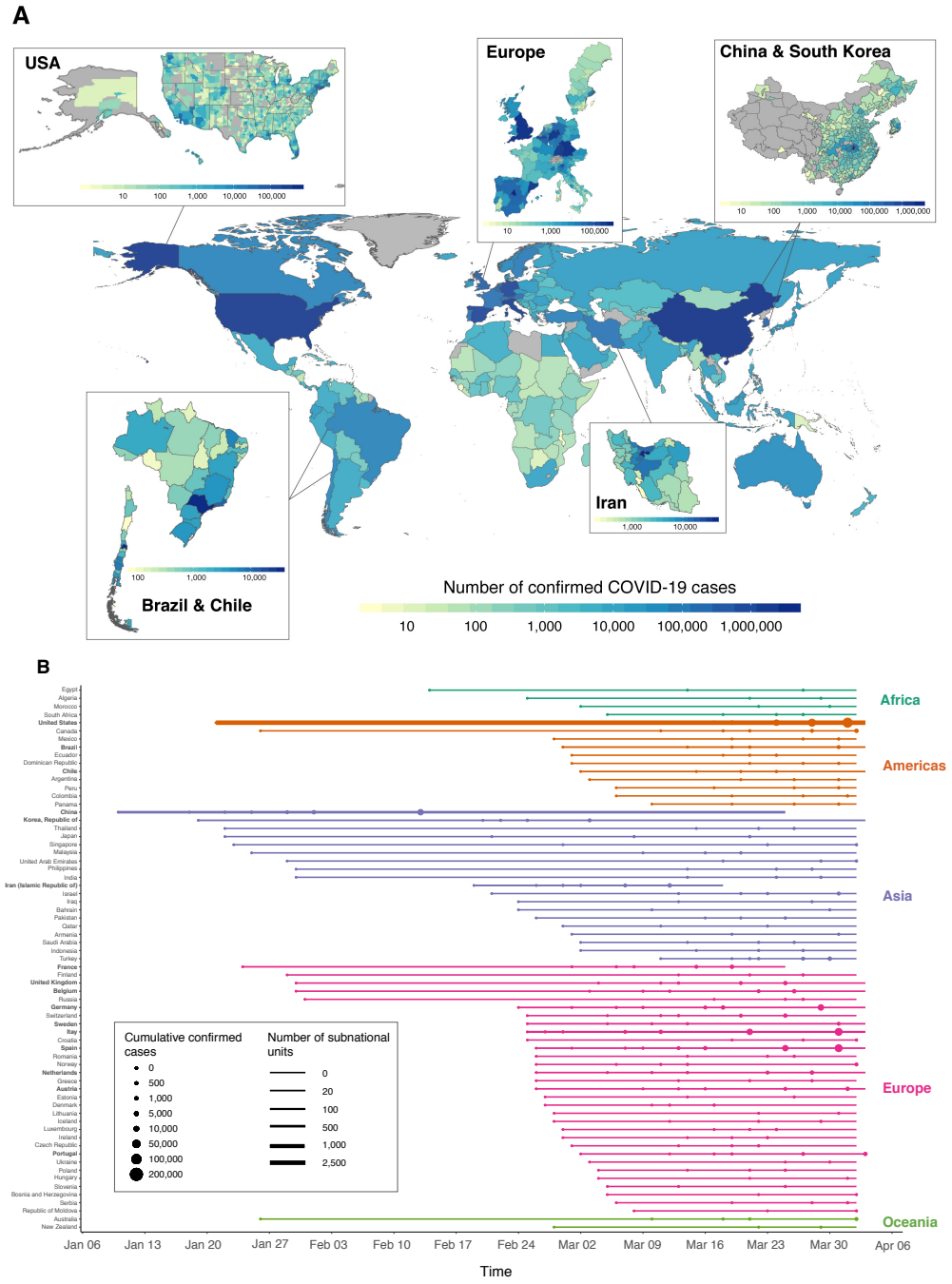

Figure 1: Global assemblage of national and subnational COVID-19 records. Panel **A** shows total confirmed cases of COVID-19 across 3,235 national and subnational units covering 173 countries. Darker colors indicate a higher cumulative number of confirmed cases as of April 10, 2020 gray indicates that no data are available. Subnational COVID-19 records were obtained for the United States, Brazil, Chile, Iran, China, South Korea, and 10 European countries. Each box shows within-country heterogeneity in COVID-19 cases for countries with subnational records. Panel **B** shows, for the subset of countries with at least 1,000 reported cumulative cases, the period of available COVID-19 data. Data from countries that are in bold text are available at the subnational level, with the number of administrative units indicated by the thickness of the time series line. Circles indicate the date when cumulative confirm cases reach specific thresholds, with larger circles indicating higher case counts.

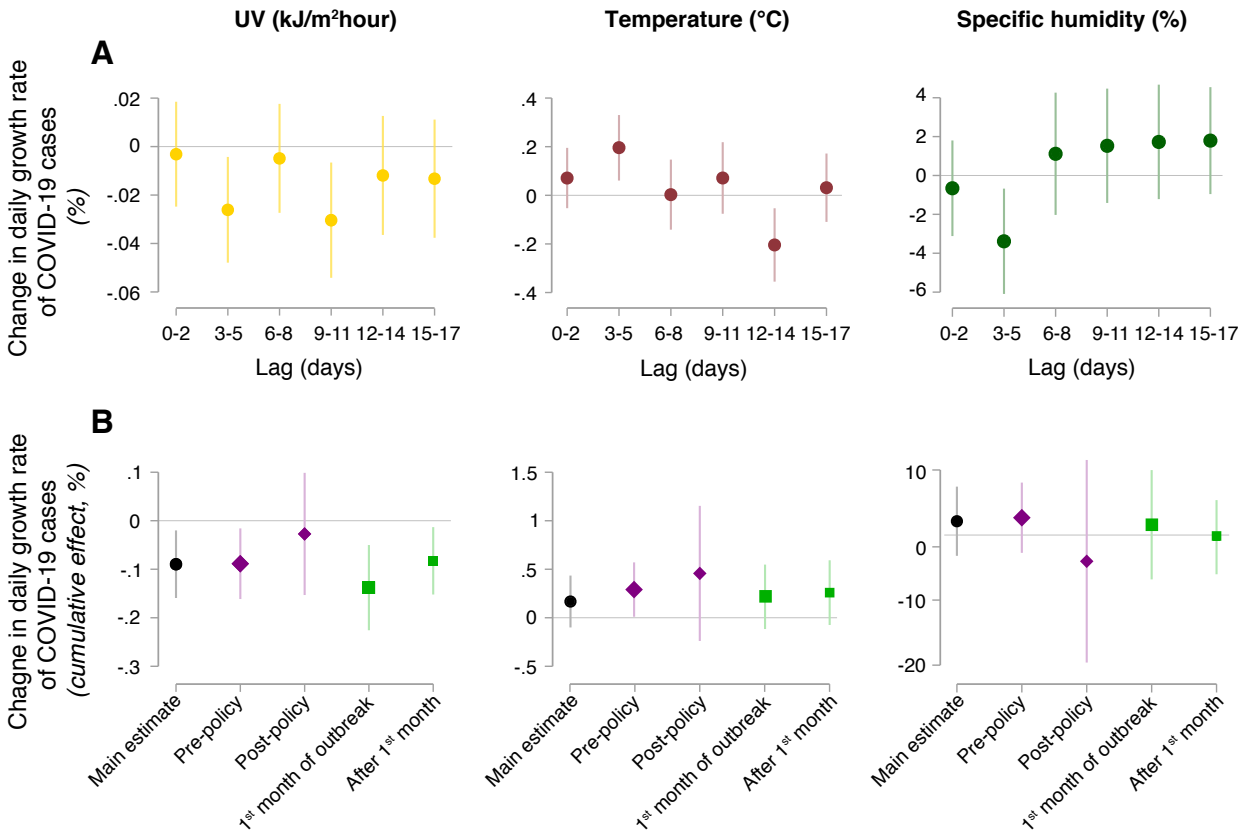

Figure 2: Empirical estimates of the dynamic relationship between COVID-19 and local environmental conditions. Panel **A** shows the estimated response of the daily growth rate (in percentage points) in confirmed COVID-19 cases as a function of lagged 3-day averaged UV radiation (first column), temperature (second column), and specific humidity (third column), occurring 0 to 2, 3 to 5, 6 to 8, 9 to 11, 12 to 14, and 15 to 17 days prior. All coefficients were estimated jointly in a statistical model leveraging a rich set of semi-parametric controls to isolate idiosyncratic variation in each weather variable (Supplementary Information Section A.2). Daily precipitation is included as a control (Fig. S6). Point estimates are indicated by circles. Vertical lines indicate 95% confidence intervals, based on standard errors that allow for arbitrary correlation of model errors within administrative units over time. Panel **B** shows the cumulative effect of all lagged coefficients for each weather variable in black, with columns corresponding to the order of weather variables in Panel **A**. In purple, treatment effects of UV, temperature, and specific humidity are reported for the period of time before an administrative unit imposed any social distancing measures (large purple diamond), and after such measures were put in place (small purple diamond). Similarly, in green, treatment effects of each weather variable are reported for the first 30 days of the location-specific outbreak (large green square), and for all dates after the first 30 days (small green square). Vertical lines indicate 95% confidence intervals. Effects of social distancing policies and outbreak duration on individual lag coefficients for all three weather variables are shown in Fig. S9.

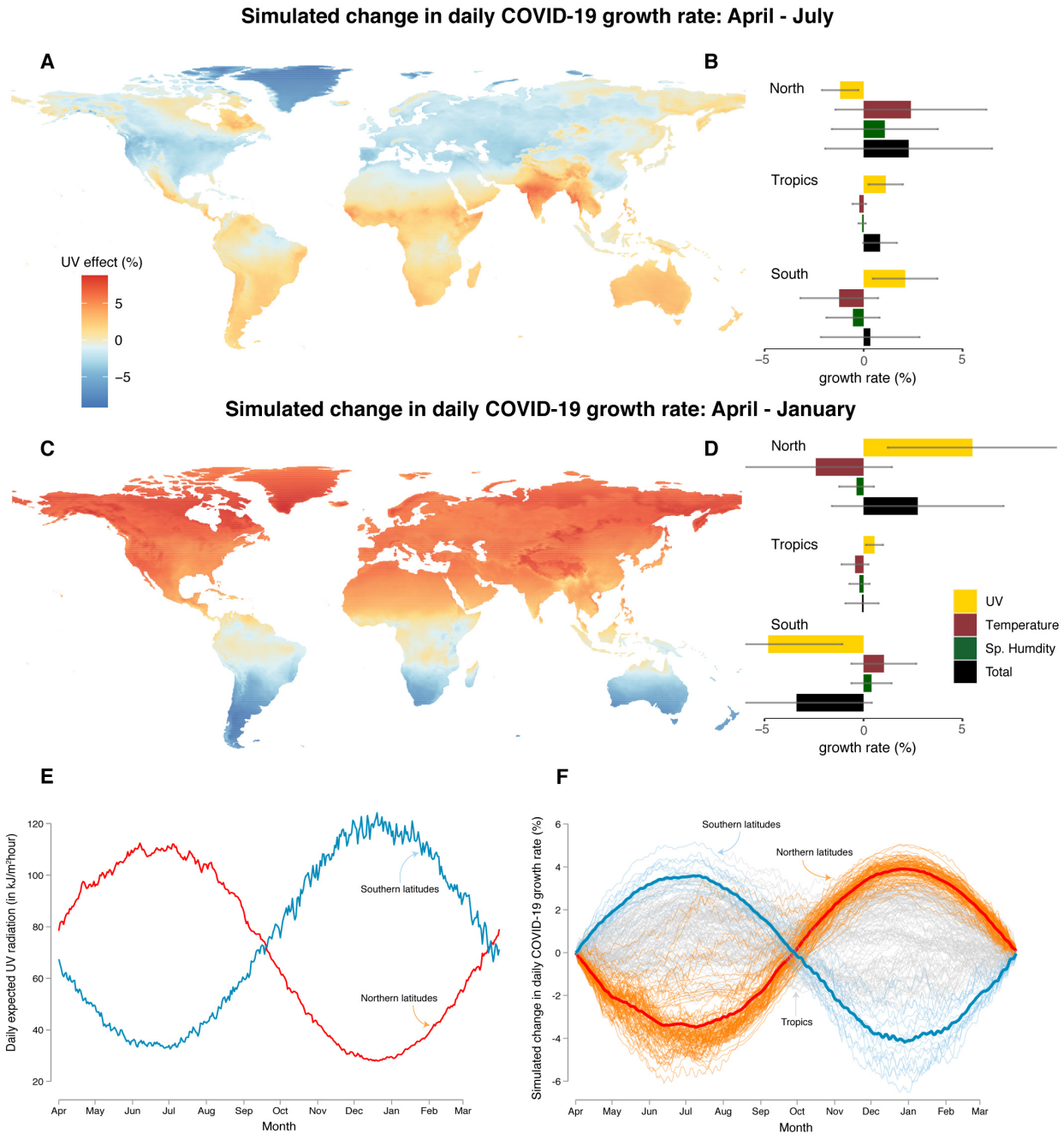

Figure 3: Simulated seasonality in the daily COVID-19 growth rate. Maps show the influence of expected seasonal changes in UV alone on the COVID-19 growth rate from **A** April to July and **C** April to January. Barcharts show the individual impacts of seasonal changes in UV (gold), temperature (maroon), and specific humidity (green), as well as their combined effect, again from April to July **B** and April to January **D**. Barcharts indicate average simulated impacts for the northern latitudes (above 23°C north), the tropics (23° south to 23° north), and the southern latitudes (below 23° south). Whiskers show 95% confidence intervals, which account only for statistical parameter uncertainty. Panel **E** shows the seasonal cycle in UV exposure for northern and southern latitudes, respectively. In panel **F**, each thin line shows a single country's simulated percentage point change in the daily growth rate of COVID-2019 cases due to changes in UV exposure between April and each day in the subsequent year for northern latitude countries (red), southern latitude countries (blue) and tropical countries (gray). Thick lines indicate average projected percentage point changes within countries in the northern (red) and southern (blue) latitudinal groups.

Methods Summary

To construct a harmonized global dataset of geolocated daily confirmed COVID-19 cases, we assemble publicly available data from national governments, subnational authorities, and newspapers. Subnational sources are described in Table S2; for all countries for which subnational records were not publicly available at the time of writing, we use national-level records provided by the Johns Hopkins University Center for Systems Science and Engineering [46]. We link COVID-19 case data to gridded daily weather data from the ERA5 reanalysis product made available by the European Centre for Medium-Range Weather Forecast [30] by calculating the population-weighted average UV, temperature, specific humidity, and precipitation for each day across all grid cells within each administrative unit (Supplementary Information Section B). We collect data on the intensity of COVID-19 testing and on social distancing policies from ref. [47] and the Oxford COVID-19 Government Response Tracker (OxCGRT) [48]. No statistical methods were used to predetermine sample size.

We statistically estimate the effect of weather on the daily growth rate of confirmed COVID-19 cases using a longitudinal (i.e. panel) regression model. Daily COVID-19 growth rates are estimated as a linear function of UV, temperature, specific humidity, and precipitation exposure over the preceding 17 days in a model allowing the effect of each environmental variable to differ across lag intervals of 3 days (Eq. S1). We include indicator variables (i.e. fixed effects) for each subnational or national administrative unit, for each day of the sample, and for each country-by-week in the sample. We calculate standard errors accounting for serial correlation across days within each administrative unit. We test for heterogeneity in the estimated effect of weather conditions on COVID-19 growth rates by interacting the lagged weather variables with binary variables indicating whether containment policies are in place and whether the observation is in the first month of the outbreak (Supplementary Information Section A.2).

Seasonal simulations (Fig. 3) use the daily seasonal climatology of UV, temperature, and specific humidity, which we calculate by averaging daily data from the ERA5 reanalysis product over the years 2015 to 2019 (Supplementary Information Section A.3).

Code availability. Code to replicate the analysis is available at: https://doi.org/10.5281/zenodo.3829621 (DOI 10.5281/zenodo.3829621)

Data availability. All data used in this analysis are compiled from free, publicly available sources. Data for replication is available at: https://doi.org/10.5281/zenodo.3829621 (DOI 10.5281/zenodo.3829621)

References

[1] Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).

[2] Sun, P. *et al.* Clinical characteristics of hospitalized patients with SARS-CoV-2 infection: A single arm meta-analysis. *Journal of Medical Virology* **92** (2020). URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.25735. https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmv.25735.

[3] Cohen, J. Sick time. *Science* **367**, 1294–1297 (2020). URL https://science.sciencemag.org/content/367/6484/1294. https://science.sciencemag.org/content/367/6484/1294.full.pdf.

[4] Lowen, A. C., Mubareka, S., Steel, J. & Palese, P. Influenza virus transmission is dependent on relative humidity and temperature. *PLoS Pathog* **3**, e151 (2007).

[5] Steel, J., Palese, P. & Lowen, A. C. Transmission of a 2009 pandemic influenza virus shows a sensitivity to temperature and humidity similar to that of an H3N2 seasonal strain. *Journal of Virology* **85**, 1400–1402 (2011). URL https://jvi.asm.org/content/85/3/1400. https://jvi.asm.org/content/85/3/1400.full.pdf.

[6] Barreca, A. I. & Shimshack, J. P. Absolute humidity, temperature, and influenza mortality: 30 years of county-level evidence from the United States. *American Journal of Epidemiology* **176**, S114–S122 (2012).

[7] Jaakkola, K. *et al.* Decline in temperature and humidity increases the occurrence of influenza in cold climate. *Environmental Health* **13**, 22 (2014).

[8] Reich, N. G. *et al.* A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States. *Proceedings of the National Academy of Sciences* **116**, 3146–3154 (2019). URL https://www.pnas.org/content/116/8/3146. https://www.pnas.org/content/116/8/3146.full.pdf.

[9] Duan, S. *et al.* Stability of SARS coronavirus in human specimens and environment and its sensitivity to heating and UV irradiation. *Biomedical and Environmental Sciences: BES* **16**, 246–255 (2003).

[10] Darnell, M. E., Subbarao, K., Feinstone, S. M. & Taylor, D. R. Inactivation of the coronavirus that induces severe acute respiratory syndrome, SARS-CoV. *Journal of Virological Methods* **121**, 85–91 (2004).

[11] Ianevski, A. *et al.* Low temperature and low UV indexes correlated with peaks of influenza virus activity in Northern Europe during 2010–2018. *Viruses* **11**, 207 (2019).

[12] Fischer, R. *et al.* Assessment of N95 respirator decontamination and re-use for SARS-CoV-2. *medRxiv* (2020). URL https://www.medrxiv.org/content/early/2020/04/15/2020.04.11.20062018. https://www.medrxiv.org/content/early/2020/04/15/2020.04.11.20062018.full.pdf.

[13] Karapiperis, C., Kouklis, P., Papastratos, S., Chasapi, A. & Ouzounis, C. Assessment for the seasonality of COVID-19 should focus on ultraviolet radiation and not 'warmer days' (2020).

[14] Wang, J., Tang, K., Feng, K. & Lv, W. High temperature and high humidity reduce the transmission of COVID-19. *Available at SSRN 3551767* (2020).

[15] Chen, B. *et al.* Roles of meteorological conditions in COVID-19 transmission on a worldwide scale. *medRxiv* (2020). URL https://www.medrxiv.org/content/early/2020/03/20/2020.03.16.20037168. https://www.medrxiv.org/content/early/2020/03/20/2020.03.16.20037168.full.pdf.

[16] Shi, P. *et al.* The impact of temperature and absolute humidity on the coronavirus disease 2019 (COVID-19) outbreak - evidence from China. *medRxiv* (2020). URL https://www.medrxiv.org/content/early/2020/03/24/2020.03.22.20038919. https://www.medrxiv.org/content/early/2020/03/24/2020.03.22.20038919.full.pdf.

[17] Bukhari, Q. & Jameel, Y. Will coronavirus pandemic diminish by summer? *Available at SSRN 3556998* (2020).

[18] Araujo, M. B. & Naimi, B. Spread of SARS-CoV-2 coronavirus likely to be constrained by climate. *medRxiv* (2020). URL https://www.medrxiv.org/content/early/2020/03/16/2020.03.12.20034728. https://www.medrxiv.org/content/early/2020/03/16/2020.03.12.20034728.full.pdf.

[19] Deschênes, O. & Greenstone, M. The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather. *American Economic Review* **97**, 354–385 (2007). URL http://www.aeaweb.org/articles?id=10.1257/aer.97.1.354.

[20] Dell, M., Jones, B. F. & Olken, B. A. What do we learn from the weather? The new climate-economy literature. *Journal of Economic Literature* **52**, 740–98 (2014). URL http://www.aeaweb.org/articles?id=10.1257/jel.52.3.740.

[21] Carleton, T. A. & Hsiang, S. M. Social and economic impacts of climate. *Science* **353**, aad9837 (2016).

[22] Sachs, J. D. & Warner, A. M. Fundamental sources of long-run growth. *American Economic Review* **87**, 184–188 (1997).

[23] National Academies of Sciences, E. & Medicine. *Rapid expert consultation on SARS-CoV-2 survival in relation to temperature and humidity and potential for seasonality for the COVID-19 pandemic (April 7, 2020)* (The National Academies Press, Washington, DC, 2020). URL https://www.nap.edu/catalog/25771/rapid-expert-consultation-on-sars-cov-2-survival-in-relation-to-temperature-and-humidity-and-poten

[24] Lauer, S. A. *et al.* The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Annals of Internal Medicine* (2020).

[25] Liu, Y., Gayle, A. A., Wilder-Smith, A. & Rocklöv, J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of Travel Medicine* (2020).

[26] Li, Q. *et al.* Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine* (2020).

[27] Kucharski, A. J. *et al.* Early dynamics of transmission and control of COVID-19: A mathematical modelling study. *The Lancet Infectious Diseases* (2020).

[28] Liautaud, P., Huybers, P. & Santillana, M. Fever and mobility data indicate social distancing has reduced incidence of communicable disease in the United States (2020). 2004.09911.

[29] Kraemer, M. U. *et al.* The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* (2020).

[30] Copernicus Climate Change Service. ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate ("2017"). URL https://cds.climate.copernicus.eu/cdsapp#!/home. Date of access: 2020-03-20.

[31] Bright, E. A., P. R. Coleman, A. N. Rose, and M. L. Urban. LandScan 2011 ("2012"). URL web.ornl.gov/sci/landscan/index.shtml.

[32] Britton, T. & Pardoux, E. *Stochastic Epidemic Models with Inference* (Springer, Cham, Switzerland, 2019).

[33] Backer, J. A., Klinkenberg, D. & Wallinga, J. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 january 2020. *Eurosurveillance* **25**, 2000062 (2020).

[34] Sun, K., Chen, J. & Viboud, C. Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: A population-level observational study. *The Lancet Digital Health* (2020).

[35] McKinnon, K. A., Stine, A. R. & Huybers, P. The spatial structure of the annual cycle in surface temperature: Amplitude, phase, and lagrangian history. *Journal of Climate* **26**, 7852–7862 (2013).

[36] Kissler, S. M., Tedijanto, C., Goldstein, E., Grad, Y. H. & Lipsitch, M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* (2020).

[37] Fernandez, M. O. *et al.* Assessing the airborne survival of bacteria in populations of aerosol droplets with a novel technology. *Journal of the Royal Society Interface* **16**, 20180779 (2019).

[38] Li, R. *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science* (2020).

[39] Kaashoek, J. & Santillana, M. COVID-19 positive cases, evidence on the time evolution of the epidemic or an indicator of local testing capabilities? A case study in the United States. *arXiv* (2020).

[40] Setti, L. *et al.* SARS-CoV-2 RNA found on particulate matter of Bergamo in Northern Italy: First preliminary evidence. *medRxiv* (2020).

[41] Ogen, Y. Assessing nitrogen dioxide (NO2) levels as a contributing factor to the coronavirus (COVID-19) fatality rate. *Science of The Total Environment* 138605 (2020).

[42] Conticini, E., Frediani, B. & Caro, D. Can atmospheric pollution be considered a co-factor in extremely high level of SARS-CoV-2 lethality in Northern Italy? *Environmental Pollution* 114465 (2020).

[43] Wu, X., Nethery, R. C., Sabath, B. M., Braun, D. & Dominici, F. Exposure to air pollution and COVID-19 mortality in the United States. *medRxiv* (2020).

[44] Gao, W., Schmoldt, D. & Slusser, J. R. *UV Radiation in Global Climate Change* (Springer, 2010).

[45] Graff Zivin, J. & Neidell, M. Temperature and the allocation of time: Implications for climate change. *Journal of Labor Economics* **32**, 1–26 (2014).

[46] Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases* (2020).

[47] Hsiang, S. *et al.* The effect of large-scale anti-contagion policies on the coronavirus (COVID-19) pandemic. *medRxiv* (2020).

[48] Hale, T., Petherick, A., Phillips, T. & Webster, S. Variation in government responses to COVID-19: version 4.0. *Blavatnik School of Government Working Paper* (2020).

[49] Ma, J. Estimating epidemic exponential growth rate and basic reproduction number. *Infectious Disease Modelling* **5**, 129–141 (2020).

[50] Hsiang, S. Climate econometrics. *Annual Review of Resource Economics* **8**, 43–75 (2016).

[51] Cameron, A. C. & Trivedi, P. K. *Regression analysis of count data*, vol. 53 (Cambridge university press, 2013).

Supplementary Information

Ultraviolet radiation decreases COVID-19 growth rates: Global causal estimates and seasonal implications

Tamma Carleton[*], Jules Cornetet[†], Peter Huybers[‡], Kyle C. Meng[§], and Jonathan Proctor[¶]

This file includes:

- Methods
- Data
- Fig. S1-S13
- Table S1-S3

[*]University of Chicago, Dept. of Economics and Energy Policy Institute (email: tcarleton@uchicago.edu)
[†]École Normale Supérieure Paris-Saclay, Département de Sciences Sociales (email: jules.cornetet@ens-paris-saclay.fr).
[‡]Huybers: Department of Earth and Planetary Sciences, Harvard University (email: phuybers@fas.harvard.edu.
[§]Bren School of Environmental Science and Management, Dept. of Economics, and Environmental Markets Solutions Lab (emLab), UC Santa Barbara and National Bureau of Economic Research (email: kmeng@bren.ucsb.edu)
[¶]Center for the Environment and Data Science Initiative, Harvard University (email: jproctor1@fas.harvard.edu)

A Methods

A.1 SEIR model

This section conducts a set of simulations using a standard susceptible (S), exposed (E), infectious (I), and recovered (R) (i.e., SEIR) model of infectious disease to guide our statistical approach. In particular, these SEIR simulations: 1) inform the use of a temporal distributed lag statistical model to capture delayed effects of weather-induced shocks to transmission on the COVID-19 growth rate, 2) demonstrate that the cumulative effect from such a distributed lag model likely underestimates the effect of environmental conditions on COVID-19 transmission, and 3) illustrate that cumulative effects on COVID-19 growth rates identified off of high frequency daily variation in environmental conditions are appropriate for simulating the effect of lower frequency seasonal variation.

The SEIR model with constant rates of exposure (σ) and recovery (γ), but time-variable transmission ($\beta(t)$) is written,

$$\begin{aligned} S'(t) &= -\beta(t)I(t)S(t), \\ E'(t) &= \beta(t)I(t)S(t) - \sigma E(t), \\ I'(t) &= \sigma E(t) - \gamma I(t), \\ R'(t) &= \gamma I(t). \end{aligned}$$

S represents the susceptible population, E the exposed population, I the infectious population, and R the recovered population. Throughout these simulations we let $\sigma = \frac{1}{4}$, $\gamma = \frac{1}{20}$ and $\zeta = \frac{1}{50}$ with units of $\frac{1}{day}$.[1] While we do not observe E, I or R directly in the data, we do observe the number of confirmed positive cases. To capture the fact that testing is not universal in any country currently, we model the evolution of these confirmed cases (C) as, $C'(t) = \zeta I(t)$, allowing a portion ζ of the infectious population to be tested each time period. The dynamics of C are proportional to those of R, such that $C = \frac{\zeta}{\gamma}R$, assuming that both variables have equal initial conditions (both are initially set to 0 in these simulations). Conceptually, this formulation allows for the possibility that a single patient tests positive multiple times, which is consistent with some reports.[2] While in reality many factors may influence the time-variable nature of transmission, $\beta(t)$, to highlight the role of weather, here we consider changing environmental conditions as the single cause of changes in transmission.

To gain intuition for how shocks to β influence the growth rate of C, defined as $\lambda_t^C = ln(C_t) - ln(C_{t-1})$, we first simulate the evolution of the disease in the SEIR model deterministically using the semi-implicit Euler method. We let transmission vary over time with linear disturbances due to changes in weather, U: $\beta(t) = \beta_0 + \beta_1 U_t$. We parameterize U_t to generate a table-top perturbation, equal to zero except for a single time step equaling one, leading to a proportional change in β equal to β_1. Relative to a control run with constant β, the growth rate in I, λ^I, is seen to rapidly rise and then undergo a quasi-exponential decay (Fig S11A). The response of the growth rate in C to changes in β is lagged and smoothed relative to that of I.

To gain intuition for this relationship analytically, we examine how growth rates in the disease-free equilibrium change as β changes. In the disease-free equilibrium (i.e. $S \approx 1$), the populations E, I, and

[1] Qualitative results are robust to substantial changes in model parameters.

[2] https://www.reuters.com/article/us-health-coronavirus-who/who-says-looking-into-reports-of-some-covid-patients-testing-positive-il=0

C grow at the same asymptotic rate, equal to $\lambda = \frac{-(\sigma+\gamma)+\sqrt{(\sigma-\gamma)^2+4\sigma\beta}}{2}$ [49] (Fig. S11). Again, assuming a linear relationship between β and U and differentiating λ^I with respect to U gives: $\frac{\partial\lambda^I}{\partial U} = \frac{\partial\beta}{\partial U}\frac{1}{\eta}$ where $\eta = (1 + \frac{\gamma}{\sigma} + \frac{2\lambda_I}{\sigma}) > 1$. This shows that an equilibrium change in β causes a smaller equilibrium change in λ, and that the degree of damping is a function of the model parameters. We are unaware of an analytical solution for the growth rate of I, or C, under a time-variable β, and note that obtaining such a solution would be useful for purposes of optimizing inferences from changes in growth rate.

To inform how a distributed lag statistical model captures delayed effects of weather-induced shocks to transmission on the case growth rate, we use a stochastic version of the SEIR model [32]. The transmission rate β is again represented as a linear function of U, but U_t is now parameterized as the sum of a sinusoid in t and Gaussian noise. Similar results are found prescribing U_t to evolve following an autoregressive moving average model with a Gaussian innovation distribution. Pooling observations from an ensemble of 300 runs, each indexed by r, we empirically estimate the effect of U on λ^I and λ^C using a distributed lag regression model ($\lambda^C_{r,t} = c_0 + \sum_{l=\ell}^{25} \alpha_\ell U_{r,t-\ell} + \epsilon_{r,t}$). Here, c_o in an intercept and $\epsilon_{r,t}$ the error term. We find that the estimated lagged effects of U on λ^C and λ^I are similar in shape and sign to the dynamic pattern in λ^C and λ^I obtained using the deterministic model (Fig. S11A,B).

The sum across lagged coefficients, $\alpha = \sum_\ell \alpha_\ell$, from the distributed lag model is referred to as the *cumulative effect*. The value of α represents the sensitivity of λ^C to a mean change in U after 25 days. For the SEIR parameters and variations in β that we specify, we empirically find that α is approximately 1/3 the value of β_1 (Fig. S12), in qualitative agreement with the expected smaller sensitivity from consideration of the equilibrium solution. Running the simulation and regression with different frequency perturbations to β results in only small changes in α. This stability supports the application of our empirically determined α from the actual data to simulate longer-term seasonal effects.

Note that the shapes of the lagged responses seen in both the stochastic and deterministic models are determined by the assumptions of the SEIR model (e.g. an exponentially distributed infectious period). Given that the dynamics of COVID-19 is unlikely to satisfy these assumptions, we should not expect lagged responses recovered from the data to precisely match these simulated responses. Further, while the primary mechanism through which weather is thought to impact COVID-19 growth is through changes in transmission, which motivates these simulations, it is possible that weather also impacts the testing rate, recovery rate, or incubation period. Thus, the estimated impacts of weather on λ^C should be interpreted as the combined effect of potentially multiple channels – both biological and social.

A.2 Statistical model

The SEIR simulations in Section A.1 suggest that a distributed lag regression model can capture delayed effects of weather-induced shocks to transmission on the COVID-19 growth rate. Implementing such a statistical model on actual data, however, requires additional consideration of the various potential confounding factors that can influence the true data generating process, but which were not included in the simple SEIR model we examine. In general, there are two challenges to causal estimation in this setting. First, surface weather conditions vary systematically as one moves away from the equator towards higher latitude locations. For example, temperatures and specific humidity both decline at higher latitudes. Because similar latitude-dependent gradients exist for other potentially relevant environmental conditions like natural disaster exposure and socio-economic indicators like GDP, a cross-sectional analysis of local mean climate conditions and COVID-19 infection rates may be biased by such confounding factors. Second, for a

given location, surface temperatures, UV exposure, humidity and rainfall all generally trend over the course of a calendar year. Because COVID-19 infection rates trend as well, such temporal dependence may also confound empirically estimated weather effects on COVID-19 with other gradually evolving determinants of infection.

This study takes a quasi-experimental statistical approach that removes such potentially confounding factors in order to isolate random variation across a set of environmental conditions: UV, temperature, humidity and precipitation. This "reduced-form" empirical approach is agnostic regarding the mechanisms through which climate variables govern the growth rate of cases, but by providing plausibly causal estimates of the role that each plays in the evolution of the virus, allows one to make counterfactual simulations of future conditions under alternative environmental conditions. Furthermore, such estimates provide empirical grounding for the parameters of more process-based models like the SEIR model.

Specifically, we estimate a longitudinal (i.e. panel) regression model using daily confirmed COVID-19 cases from 173 countries from January 01, 2020 to April 10, 2020. Our outcome of interest is the growth rate of cumulative COVID-19 cases in administrative (i.e., national/subnational) unit i between day t and $t-1$, $\lambda_{it}^{C} = \ln C_{it} - \ln C_{it-1}$. Because of the delay between initial COVID-19 exposure and confirmed detection (Section A.1), we model the growth rate in cumulative COVID-19 cases using the following distributed lag model

$$\lambda_{it}^{C} = \sum_{\ell=0}^{\ell=L} \alpha_{\ell}^{UV} UV_{i,t-\ell} + \sum_{\ell=0}^{\ell=L} \alpha_{\ell}^{T} T_{i,t-\ell} + \sum_{\ell=0}^{\ell=L} \alpha_{\ell}^{H} H_{i,t-\ell} + \sum_{\ell=0}^{\ell=L} \alpha_{\ell}^{P} P_{i,t-\ell} + \theta' \mathbb{Z}_{it} + \epsilon_{it} \tag{S1}$$

where for administrative unit i and day t, $UV_{i,t-\ell}$, $T_{i,t-\ell}$, $H_{i,t-\ell}$, and $P_{i,t-\ell}$ are population-weighted daily average UV (in kJ/m^{2}hour), temperature (in degrees centigrade), specific humidity (in %), and precipitation (in mm), respectively, observed ℓ days ago. In robustness checks (Fig. S7), we show that specific and relative humidity generate very similar results. Because daily weather variables observed in a given location are highly serially-correlated, it is often difficult to discern a statistically precise time pattern of lagged effects when including lagged daily UV, temperature, humidity, and precipitation. To reduce this noise, we construct lagged weather terms that are averaged over 3-day periods, such that the lag index is now in terms of 3-day averages, $\ell \in \{0\text{-}2, 3\text{-}5, 6\text{-}8, 9\text{-}11, 12\text{-}14, 15\text{-}17\}$. In a robustness check, we also include leads of 3-day average values for UV, temperature, humidity and precipitation.

To isolate plausibly random variation in weather conditions [19, 50], we include $\mathbb{Z}_{it}$, a vector of semi-parametric controls. In our baseline specification, $\mathbb{Z}_{it}$ includes a full set of national/subnational unit-specific dummies, which remove any time-invariant differences in growth rates of COVID-19 cases and environmental variables across administrative units. These spatial "fixed effects" address the concern that baseline population characteristics (e.g. economic activity, population density) may be correlated both with COVID-19 infection rates and with average weather conditions. Second, $\mathbb{Z}_{it}$ includes day-specific dummies to remove any common global determinants of COVID-19 growth rates. These temporal fixed effects account for global daily circumstances that may influence COVID-19 growth rates such as WHO's declaration of COVID-19 as a global pandemic. To account for local trends in both COVID-19 growth rates and weather during this period, $\mathbb{Z}_{it}$ includes country-by-week dummies, which flexibly account for country-specific temporal trends and shocks in COVID-19 growth rates and weather. Importantly, these dummy variables capture gradual seasonality occurring differentially across the globe as COVID-19 evolves; any trending variables correlated with seasonal cycles and with the evolution of the epidemic within a country are thus accounted for. The influence of adding this suite of controls on the residual variation in UV, temperature, and COVID-19 growth

rates is shown visually for two selected regions in Fig. S3B. Finally, we cluster standard errors, ϵ_{it} at the administrative level which allows for data-driven heteroskedasticity and serial correlation in the error terms of each administrative unit. Clustering at the country level does not discernibly change the precision of our estimates.

The results of several robustness checks are given in Table S1 wherein a suite of controls are alternatively examined. Examined controls include the number of days since the initial outbreak of COVID-19 in each location (col. 1), country-specific linear trends (col. 2), country-by-week fixed effects (col. 3, and in combination with others in cols. 4-6), a placebo "lead" weather variable measuring future exposure (col. 4), timing of policies such as school closures (col. 5), stringency of the COVID-19 testing regime at national level (col. 6), increased spatial resolution of our week-specific dummies to include subnational administrative unit-by-week fixed effects (col. 7), and use of country-by-day fixed effects, as opposed to country-by-week, (col. 8), which requires dropping all data from countries without subnational COVID-19 records, as daily weather variables are collinear with these dummy variables.

To control for social distancing policies (Table S1, col. 5), which vary across space and time, we add to the regression model in Eq. S1 a dummy variable equal to 1 when any one of three policies are in place: school closures, work from home ordinances, and event cancellations (Section B.3). To control for changes over space and time in the degree of COVID-19 testing (Table S1, col. 6), we use national-level records from OxGCRT (Section B.3) that categorize each country's testing regime. Categories are: No testing policy (coded as 0), only testing those who have symptoms *and* meet specific criteria, such as being essential workers or coming into contact with a known case (coded as 1); testing of anyone showing COVID-19 symptoms (coded as 2); and open public testing, such as drive-through testing, available to asymptomatic people (coded as 3). A variable indicating which regime a country falls into on any given day is added to the regression model shown in Eq. S1. Modelling this ordinal variable as four binary variables gives nearly identical results.

Finally, we estimate a Poisson Pseudo-Maximum Likelihood estimator, in place of the ordinary least squares regression shown in Eq. S1 (col. 9, and col. 5 of Fig. S6). We do so for two reasons: first, the distribution of new cases is very skewed; second, if climate conditions operate solely through the transmission parameter β (Section A.1), changes in climatic conditions cannot lead to negative growth rate effects, as transmission cannot be negative. Because the former concern is not empirically large (in Fig. S13, we show that our residuals from estimation of Eq. S1 are approximately normally distributed), and because climate variables may influence growth rates through transmission as well as other channels, such as behavior regarding testing, we include this model as a robustness check only. The estimating equation relates new cases realized between day $t-1$ and day t, denoted ΔC_{it}, to lagged climatic exposure as follows:

$$\Delta C_{it} = exp\left(\sum_{\ell=0}^{\ell=L} \alpha_\ell^{UV} UV_{i,t-\ell} + \sum_{\ell=0}^{\ell=L} \alpha_\ell^{T} T_{i,t-\ell} + \sum_{\ell=0}^{\ell=L} \alpha_\ell^{H} H_{i,t-\ell} + \sum_{\ell=0}^{\ell=L} \alpha_\ell^{P} P_{i,t-\ell} + \theta' \mathbb{Z}_{it} + \rho C_{it-1}\right) + \epsilon_{it} \quad \text{(S2)}$$

We control for lagged cumulative cases C_{it-1} as new cases ΔC_{it} are proportional to the level of infected people in the population (Section A.1). All other variables are defined as in Eq. S1. While standard Poisson models impose that the first and second moments of the outcome be equal, we address this overdispersion issue by clustering standard errors at the administrative unit level. This adjustment relaxes the assumption of equal first and second moments by allowing arbitrary forms of within-administrative unit heteroskedasticity and serial correlation in the error term ϵ_{it} [51].

In Fig. 2B, we show cumulative effects of lagged responses of COVID-19 to UV, temperature, and specific humidity. The point estimates shown in black are the sum of the estimated coefficients across lag lengths,

$\hat{\alpha}_\ell$. We additionally show heterogeneity in this cumulative effect across policy regimes (purple diamonds) and duration of outbreak (green squares). The former coefficients are generated by estimating a version of Eq. S1 in which each weather variable lag is interacted with the corresponding lagged value of a policy dummy variable. This dummy variable is equal to 1 when any one of three policies are in place: school closures, work from home ordinances, or event cancellations (Section B.3). "Pre-policy" cumulative effects are then computed using estimated lagged effects of each weather variable when the policy dummy is set to 0; in contrast, "post-policy" cumulative effects are computed using estimated coefficients when the policy dummy is set to 1. Similarly, to recover heterogeneity by duration of outbreak, we define a dummy variable equal to 1 when an observation for a given location and day occurs at least 30 days after the first recorded COVID-19 case within that population. This dummy variable is then interacted with each lagged weather variable in a regression otherwise identical to Eq. S1. Cumulative effects for the first month of outbreak are computed using estimates of lagged weather variable effects when the outbreak duration dummy is 0; "after first month" cumulative effects are similarly computing by setting the outbreak duration dummy to 1.

A.3 Seasonal simulations

To conduct seasonal simulations, we calculate the daily seasonal climatology of UV, temperature and specific humidity by averaging daily data from the ERA5 reanalysis product over the years 2015 to 2019 (Section B.4). In Figs. 3A-D, we represent the monthly effect of each climate variable on the predicted COVID-19 growth rate as the product of the cumulative effect of each variable estimated in Eq. S1 and the average hourly weather over each calendar month. To capture differential seasonality across time, we show in Fig. 3A-D the difference between predicted growth rates under the climatology of April and under the climatology of July (A-B) and January (C-D). In Figs. 3E-F and S10, we represent country-specific daily seasonality (Fig. 3F) as the product of the cumulative effect of UV only (estimated from Eq. S1) and the average daily population-weighted UV for each country (daily UV variation is shown in Fig. 3E). Again, differences are shown relative to the April climatology.

B Data

B.1 COVID-19 case data

To statistically estimate a plausibly causal, global relationship between local weather variations and the pattern of COVID-19 transmission we construct a harmonized global dataset of geolocated daily confirmed COVID-19 cases at what we believe to be the finest spatial scale assembled to date. We use data obtained from national governments, subnational authorities and newspapers, ultimately covering 3,235 administrative units across 173 countries and five continents.

B.1.1 National-level COVID-19 data

For all countries for which subnational records were not publicly available at the time of writing, we use national data assembled by the Johns Hopkins University Center for Systems Science and Engineering [46] for the period between January 01, 2020 to April 10, 2020.[3] We omit observations from the Diamond Princess cruise ship, due to uncertain weather exposure of the passengers.

[3] Available at https://github.com/CSSEGISandData/COVID-19 and accessed via https://github.com/RamiKrispin/coronavirus.

B.1.2 Subnational COVID-19 data

Table S2 describes the characteristics of and sources for the COVID-19 case data we collected and compiled at the subnational level. Below, we provide some additional detail regarding data cleaning and manipulation for each individual country.

In most countries, we directly obtain subnational reports of the daily number of newly confirmed COVID-19 cases. To compute cumulative case counts at the daily level, we then compute cumulative sums for each subnational unit. When only cumulative COVID-19 cases are available on a daily basis, we take first differences in the time series for each subnational unit to obtain the number of new cases detected on each day. If not mentioned otherwise, we assume that missing values after the start of the epidemic in a given subnational unit correspond to zero new cases. Because we obtain subnational case data from ref. [47] for Iran and China, we follow their imputation method for addressing missing data in these two countries; details of this method are described by the authors.[4]

In many countries, additional data cleaning was required to accurately and consistently match new cases to the day on which they were detected, as opposed to the day on which they were reported. Harmonizing the data in this way reduces measurement error when estimating a common lagged response across the pooled sample. To do so, we track the dates and hours of the day on which new cases were released; when new cases are obtained from morning reports (before noon), we assign cases to the previous calendar day. Details on such corrections are presented below for each country. We compare our compiled subnational COVID-19 case data with case data reported at national level by John Hopkins University (JHU) and by the European Center for Disease Prevention and Control (ECDC) (Fig. S1).[5]

Austria (1st administrative level) See Table S2 for details. Because no alternative archived reports are available for Austria, we verify our data against data stored in the GitHub public repository "covid-19-eu-data", which provides time series for COVID-19 cases in European countries based on the scraping of official reports.[6] Our figures correspond to the official afternoon reports.

Belgium (1st administrative level) See Table S2 for details. We append data from two versions of the Wikipedia article "2020 coronavirus pandemic in Belgium." The current Wikipedia page (as of April 13, 2020) provides data starting on March 1, 2020. Data from January 30 to March 1, 2020 were webscraped from an earlier version of the same article (accessed on April 6, 2020). The distribution of cumulative cases on March 1, 2020 in the current article matches those from the previously collected time series. We drop data for April 7 and April 8, 2020, as we detect a discontinuous drop in new cases and increase in missing values. Our numbers have been verified against Sciensano data[7] for the days covered by both sources.

Brazil (1st administrative level) See Table S2 for details. We directly obtain records of daily new cases and compute time series of cumulative cases by integrating new case reports. For São Paulo, we add an additional case to the cumulative case count for February 25, 2020, based on newspaper reporting that a single case was already present.[8] We confirm that our data match the official source.

[4] Available here: https://www.dropbox.com/s/1xvskw6dark5310/2020321_GPL_COVID_appendix.pdf?dl=0.
[5] ECDC data have been directly downloaded from the ECDC website: https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide.
[6] https://github.com/covid19-eu-zh/covid19-eu-data/
[7] Available at: https://epistat.wiv-isp.be/covid/.
[8] See for instance: https://newslab.com.br/primeiro-caso-do-covid-19-e-confirmado-no-brasil/ (in Portuguese).

Chile (1st administrative level) See Table S2 for details. As stated in the daily reports, the information provided in official publications document cases reported on the previous day. In order to associate new cases with the date of detection and not the date of announcement, we correct the data webscraped from Wikipedia by lagging each date by one.

France (1st administrative level) See Table S2 for details. France overseas territories have been removed from the analysis due to the low number of cases at the time of data collection (116 cases distributed over 7 territories on March 25, 2020). On March 25, 2020, the French Public Health Agencies stopped publishing COVID-19 cases data at the regional level. Because the cumulative number of cases proposed by Wikipedia after this date is systematically below official figures at the national level and because the corresponding sources are not verifiable, we retain data only until March 25, 2020. It was not possible to find archived reports for French COVID-19 cases at the regional level for data verification purposes. We thus compare our data against the time series offered on the open platform for French public data.[9] The number of cases reported in both datasets are very similar.

Germany (1st administrative level) See Table S2 for details. While the Robert Koch Institute (RKI) publishes case data for COVID-19, these data do not exist prior to March 3, 2020. We therefore rely on webscraped data from Wikipedia, which is obtained from newspaper articles, and validate these data against those available from RKI. Between March 4 and March 10, 2020, the Wikipedia figures do not match the Robert Koch Institute (RKI) reports. Hence, we recode manually the series for these days, using the official data from RKI. Due to the relative novelty of the epidemic at that time, some reports are inconsistent with the figures presented in preceding reports. In case of such inconsistencies, we consider the most recent report as the most reliable one and correct our number of cases accordingly.

On March 17, 2020, the RKI stopped updating its data manually and switched to an automated process based on the data electronically transmitted up to 11:00pm on the previous day.[10] After this date, we correct the date in our data by lagging records by one day, in order to retrieve the accurate day of detection. For March 17, we sum new cases recovered from the reports of March 17 (confirmed cumulative infections up to March 17, 2020, 11:00pm) and March 18 (confirmed cumulative infections up to March 18, 2020, 0:01am).

Both Wikipedia and the Robert Koch Institute point out that some reports are missing, but do not consider this information in their computation of the cumulative cases series. The number of total confirmed cases reported is thus artificially stable for the dates in which reports are known to be missing. We therefore code as missing the new cases and confirmed cumulative cases for North Rhine-Westphalia on March 10 and 11, for Saxony-Anhalt on March 26 and 28 and for Baden-Wurttemberg and Hesse on March 27, to account for the absence of data collection for these dates.

Our panel data on cumulative cases begins with 14 cases in Bavaria on February 24, 2020. As it is unlikely that these 14 cases appeared all at once, we set the initial value of our new cases series as missing.

Iran (1st administrative level) See Table S2 for details. The number of new cases for all regions on March 2 and March 3, 2020, are missing, due to an absence of reporting. These missing values have been imputed following the method implemented by ref. [47], who used and verified the same source of data.

[9] https://www.data.gouv.fr/en/datasets/fr-sars-cov-2/

[10] https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/2020-03-17-en.pdf?__blob=publicationFile

Netherlands (1st administrative level) See Table S2 for details. The table we obtain from Wikipedia associates the number of new confirmed cases with the day on which they were first announced. As official reports for the Netherlands are published in the morning, we correct the date by lagging reported cases by one day, relative to that provided by Wikipedia.

Portugal (1st administrative level) See Table S2 for details. Official reports until March 9, 2020, were published in the late afternoon. On March 10, however, the Directorate-General of Health began to publish morning reports compiling the number of total confirmed cases up to midnight on the previous day. As the Wikipedia article we webscrape does not consider this change, we correct the date for each day after March 10, 2020, by lagging the case records by one day. For March 10 itself, we discard the row associated to March 10 on Wikipedia (2 new confirmed cases) and only keep the one associated to the report published on March 11.

These data have been verified against official reports. For March 2, 2020, as there was no official report at this stage of the epidemic, we verified the information in newspapers documenting the first occurrence of the epidemic in Portugal.[11]

South Korea (1st administrative level) See Table S2 for details. The time at which official counts have been released changes over the sample period. Until March 1, 2020, updates to case records often occur twice per day. The confirmed new cases announced in each report are those that have been detected since the last report: the new cases announced in the afternoon have thus been detected within the day, since the morning count. From January 30 to March 1, we group the morning count with the afternoon count of the previous day, to get a detection period covering 9:00am on the previous day to 9:00am on the current day. On these dates, the date of the afternoon count has been kept in order to match the day of detection. On March 1, we sum the cases obtained from the afternoon report to the new cases extracted from the evening report, released at midnight. From March 2 forward, the Korean Center for Disease Control (KCDC) publishes morning reports containing information for the previous day, from 00:01am to 11:59pm. As the Wikipedia article correctly considers this change, no change has been made on the date of new cases for March 2 onward.

Spain (1st administrative level) See Table S2 for details. Spanish reports are published in the morning and contain information about the previous day. We thus correct the date in the data obtained in a public GitHub repository by lagging case counts by one day in order to accurately recover the day of detection. The number of new cases for Ceuta and Melilla have been summed to match the spatial shapefile we use for aggregating gridded climate data.

Sweden (1st administrative level) See Table S2 for details. Swedish official data regarding the COVID-19 epidemic are updated daily at 11:30am. To match the day of detection, we lag the case count reported by Wikipedia by one day. The first COVID-19 case in Sweden was observed on February 3, 2020, in the Jönköpping region. As no additional cases were detected during the three weeks following the occurrence of this first case, we drop it from our continuous confirmed new cases series, which begin on February 25, 2020. However, we keep it when computing the number of cumulative cases.

[11] See for instance: https://www.reuters.com/article/us-health-coronavirus-portugal/portugal-registers-first-two-cases-of-coronavirus-sic-television-idUSKBN20P1BB.

United Kingdom (1st administrative level) See Table S2 for details. Because there were about 1700 cases that had not been precisely located within England at the time of initial data collection (April 6, 2020), we aggregated case data to the level of England, Wales, Scotland, and Northern Ireland, instead of using National Health System (NHS) regions. In the United Kingdom, confirmed new and cumulative COVID-19 cases are announced in the morning. We thus lag cases reported by Wikipedia by one day to accurately reflect date of detection. Since the date of our initial data collection, Public Health England has published complete time series at the NHS region level and at the county level. We checked our data against these series and verified that they were nearly identical.

China (2nd administrative level) See Table S2 for details. We drop 29 cities that could not be merged with climate data, based on publicly available geographic shape files. It was not possible to check the time at which city reports were issued. As a result, we consider the date of announcement as the date of detection. Missing data have been imputed following the interpolation method performed by ref. [47].

Italy (2nd administrative level) See Table S2 for details. The number of cases is updated daily at the end of the afternoon. Our data are almost identical to those obtained at national level from JHU, although JHU data display a break on March 12 due a to a delay in JHU updates, an issue reported for several countries in this dataset.[12]

United States (2nd administrative level) See Table S2 for details. For the United States, all cases are counted on the date they are first announced, and cases are located at the place where they are treated. Although the *New York Times* mostly uses the official counties as the unit of analysis, a few exceptions are worth mentioning:[13]

- The five boroughs of New York City have been gathered under the label "New York City";
- The COVID-19 cases for Cass (MO), Clay (MO), Jackson (MO) and Platte (MO) counties are exclusive of the cases detected in Kansas City, shown on their own. We drop observations under the label "Kansas City, Missouri", which does not correspond to any official county;
- All cases for Chicago are reported within Cook County (IL).

We download all the data from the *New York Times* repository, and keep only the cumulative cases, computing new cases using first differences. Because some county-level series start with strictly positive numbers (up to 37 cases on the first day), we define the first observation of each new cases series as missing, but keep this number in our cumulative cases series.

B.2 Population data

Our main outcome variable is the first difference in the natural logarithm of daily cumulative cases per 1 million people. At the national level, we use the country-level population in 2018 (the most recent year available) from the World Bank's World Development Indicators.[14] No homogeneous source of data have

[12] Up to April 17, 2020, there are more than 10 open issues on this topic on the CSSEGISandData/COVID-19 public repository managed by JHU. See for instance: https://github.com/CSSEGISandData/COVID-19/issues/619.

[13] Detailed information are provided on the COVID-19 GitHub repository of the *New York Times*.

[14] Available at http://data.worldbank.org/data-catalog/world-development-indicators

been found at the subnational level. We therefore obtain the most recent data available from each country's national office of statistics. Detailed information on each source can be found in Table S3.

B.3 Policy and COVID-19 testing data

We collect data on the intensity of COVID-19 testing and on social distancing policies from ref. [47] and the Oxford COVID-19 Government Response Tracker (OxCGRT) [48], described below. Briefly, from these data we obtain – for each administrative unit in our analysis – a policy variable, which is equal to 1 if any policy that closes schools, closes workplaces, or cancels public events is implemented, and 0 otherwise. We also obtain an ordinal variable with four levels that describes the intensity of testing.

Social distancing policy data from ref. [47] The first set of variables we use has been compiled by ref. [47] in their study on the effect of large-scale anti-contagion policies on the COVID-19 pandemic. The authors collect policy data at the subnational scale for China (2nd administrative unit), France (1st administrative unit), Iran (1st administrative unit), Italy (2nd administrative unit) and the United States (1st administrative unit). We directly merge these policy data into our database for the corresponding dates. For the United States, we match each county within a state to the state-level (i.e. 1st administrative level) data from ref. [47], as no county-level policy data are available.

We use three variables from this study, which match with variables within the OxCGRT dataset. They are defined in in ref. [47] as:

1. **school_closure**: "A policy that closes school and other educational services in that area."
2. **work_from_home**: "A policy that requires people to work remotely. This policy may also include encouraging workers to take holiday/paid time off."
3. **event_cancel**: "A policy that cancels a specific pre-scheduled large event (e.g. parade, sporting event, etc.). This is different from prohibiting all events over a certain size."

All these variables are binary variables. They take the value 1 starting on the day a policy is implemented, and 0 if the policy is not implemented. Only policies that are legally enforced are considered here; optional policies and non-binding recommendations from governments are not included.

The Oxford COVID-19 Government Response Tracker (OxCGRT) The second set of policy variables we use contains data at the national level for over one hundred countries across the world. The original dataset has been compiled by a group of researchers affiliated to the Blavatnik School of Government at Oxford [48].[15]

To match the data from ref. [47] we collect three variables from this database:

1. **School_closing**: "Record closings of schools and universities".
2. **Workplace_closing**: "Record closings of workplaces".
3. **Cancel_public_events**: "Record cancelling public events".

[15] Data can be downloaded here: https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker The working paper presenting the construction of the database can be found at https://www.bsg.ox.ac.uk/sites/default/files/2020-04/BSG-WP-2020-031-v4.0_0.pdf.

These variables were initially coded as categorical variables taking the values: not implemented, optional and legally enforced.[16] We recode these variables as binary variables to match the format of policy data from ref. [47], setting the value to 1 if the policy is legally enforced and zero otherwise.

COVID-19 testing data OxGCRT provides a measure of testing access and the testing efforts deployed by national governments. This ordinal variable takes on the following values: absence of testing policy (0); very restricted access to testing, conditional on both symptoms and past and present exposure or professional situation (1); testing only individuals with symptoms (2); and deployment of open public testing (3).

Final data manipulation Both the the data from ref. [47] and the OxCGRT data have been merged with our epidemiological dataset using the corresponding dates and administrative levels, giving preference to subnational data when available. From these data we create a binary policy variable that takes the value of 1 if any policy closing schools, workplaces or public events was enacted and 0 otherwise.

B.4 Weather data

We use the ERA5 reanalysis product from European Centre for Medium-Range Weather Forecasts (ECMWF), which provides daily gridded weather variables at the 0.25° latitude by 0.25° longitude resolution [30].[17] Specifically, for January 01, 2020 to April 10, 2020, we collect hourly downward UV radiation at the surface (in J/m^{2}hour), 2-meter temperature (in degrees centigrade), total 2-meter precipitation (in mm), and 1000 hPA specific humidity (in %). 2-meter and 1000 hPA roughly correspond to conditions near the earth's surface. In robustness checks (Figure S7), we include relative humidity, which we sample at 1000 hPA. We average UV, temperature, and specific humidity across hours in the day to obtain daily average measures, while we sum precipitation across hours in the day to obtain daily total precipitation.

We link gridded weather data to administrative-level COVID-19 cases by aggregating grid cell information over administrative (e.g. country, province, or county) boundaries. To capture climatic conditions reflective of population exposure, we average across grid cells weighting by the cross-sectional gridded distribution of population in 2011 from LandScan [31]. For example, administrative-level daily population-weighted average temperature is computed as $T_{it} = \sum_{z \in i} \omega_{gi} T_{gt}$, where g indicates grid cell, i indicates an administrative unit, and ω_{gi} is the share of administrative unit i's population that falls within grid cell g.

To estimate projected seasonal conditions and their influence on COVID-19 transmission, we construct daily gridded UV radiation, 2-meter temperature, and 1000 hPA specific humidity from ERA5, as described above, over the last five years (2015-2019). Using the average conditions for each calendar day over the past five years as a proxy for expected seasonal variation through 2020 and into early 2021 (Fig. S5), we compute daily averages across all five years of daily temperatures at both grid cell level (Fig. 3A and C) and at country level (Fig. 3D), in the latter case using the same aggregation method described above.

UV radiation is represented as including wavelengths from 200 nm to 440 nm. Limitations associated with the representation of radiative transfer associated with ERA5 reanalysis prevent us from distinguishing between UVa (400 to 315 nm), UVb (315 to 280 nm), and UVc (280 to 100 nm) effects. Higher energy UVc and UVb radiation may more rapidly deactivate SARS-CoV-2 than UVa but is also more readily absorbed in the atmosphere. A more detailed analysis of the transmission and reflection of different types of UV in association with destruction of the SARS-CoV-2 is a fruitful area for future research.

[16] The original codebook associated with this database is available at https://www.bsg.ox.ac.uk/sites/default/files/2020-04/BSG-WP-2020-031-v4.0_0.pdf.

[17] Available at https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview

C Supplementary Figures

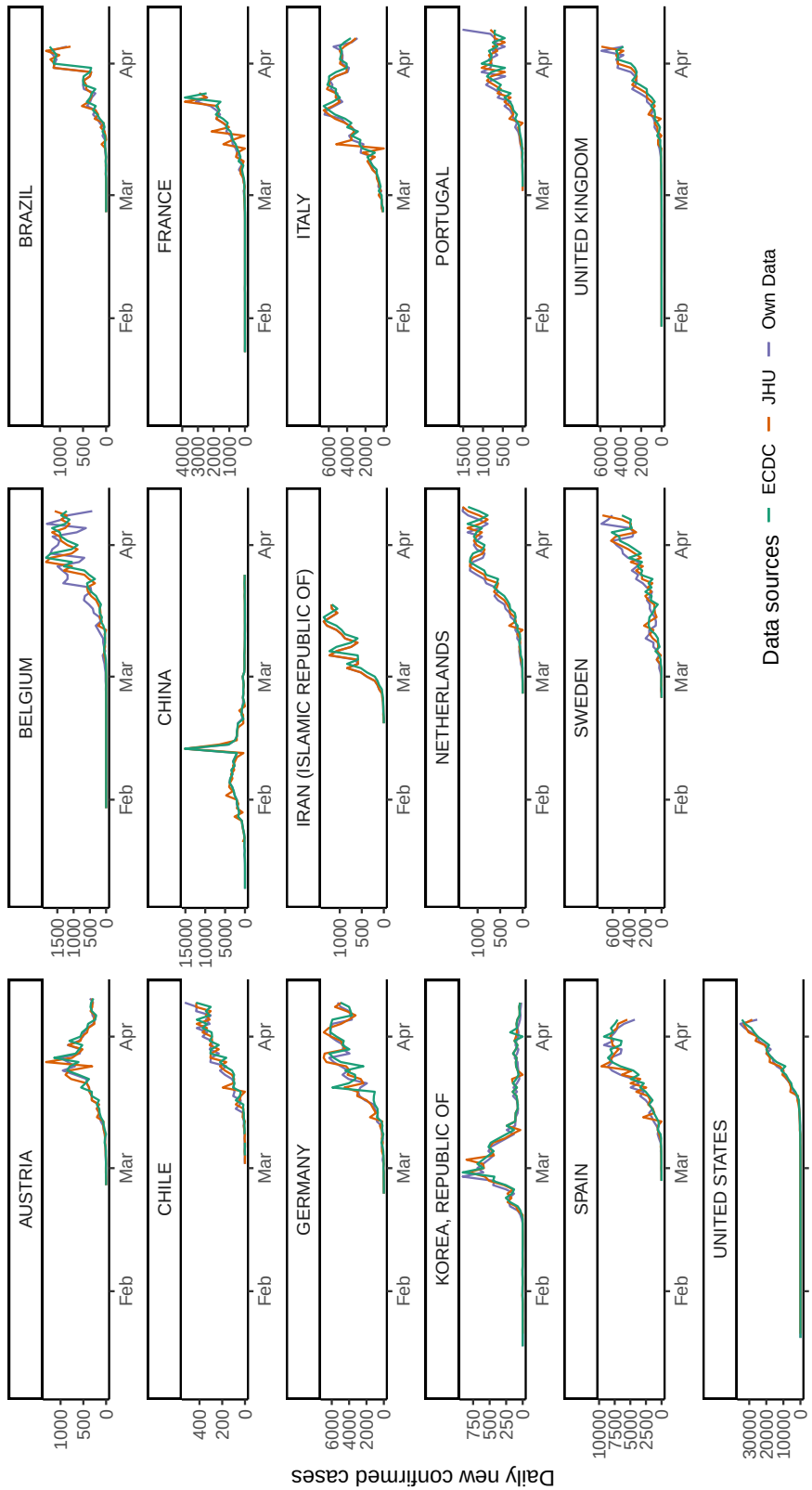

Figure S1: Comparison of newly assembled subnational data to JHU-CSSE and ECDC national-level data sources. To verify the quality of our assembled subnational COVID-19 case records, we show here the daily evolution of confirmed new cases at the national level across three distinct data srouces. First, we aggregate our subnational records to the national level; these time series are shown in purple. Second, we show in red national time series from the publicly available JHU Center for Systems Science and Engineering (JHU) data. Finally, in green we show national time series for European countries provided by the European Center for Disease Prevention and Control (ECDC). In most cases, the largest difference across datasets is due to the adjustments in dates we have made to accurately associate new cases to the date of their detection (Supplementary Materials Section B).

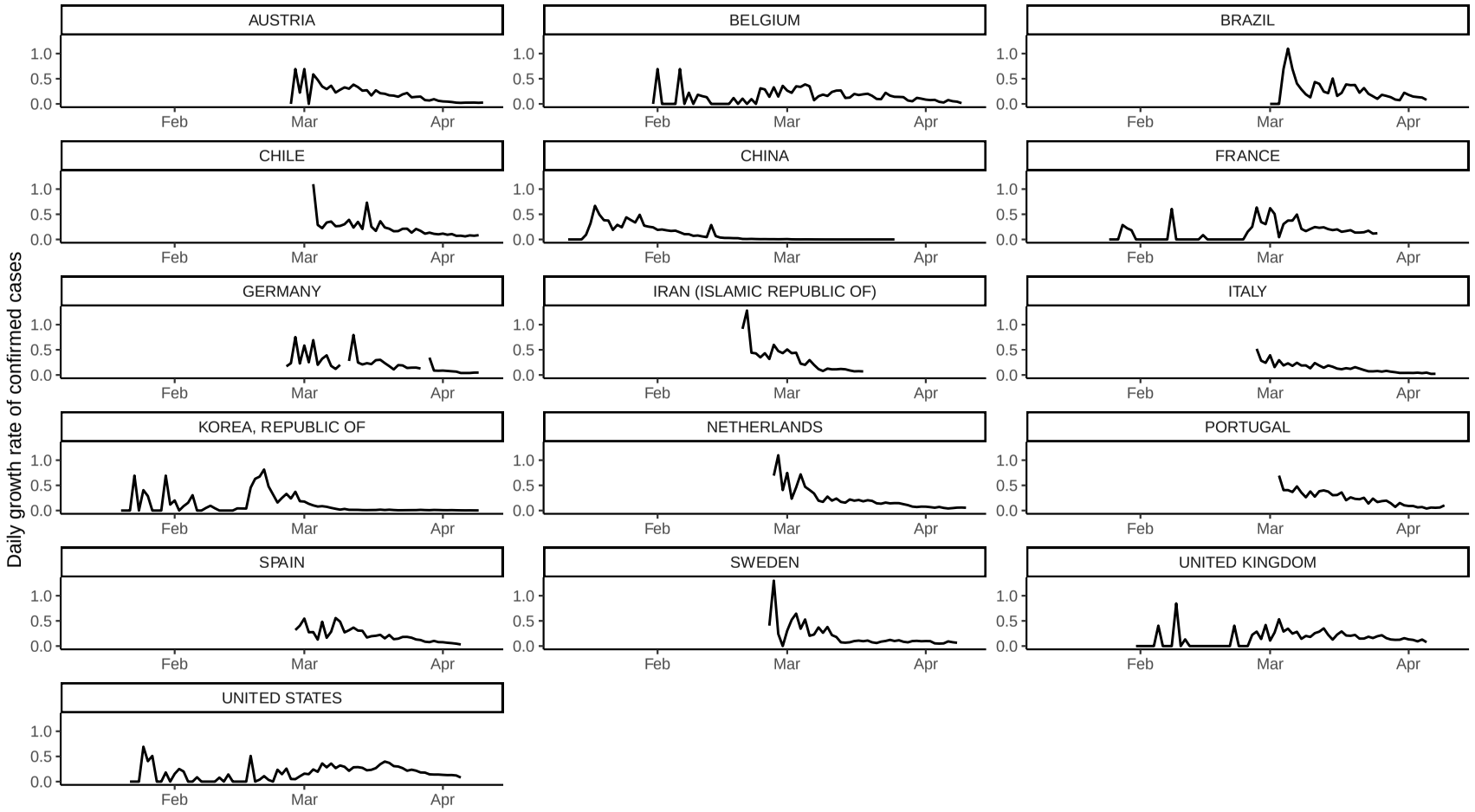

Figure S2: COVID-19 case growth rates in regions with subnational data. Here we aggregate confirmed subnational case data to the national level and plot the national COVID-19 growth rate over time from the regions where we have subnational data; subnational data from these regions comprise the majority of the sample observations used to estimate Eq. S1. Note that two missing values exist in Germany due to negative new case values, which we also drop from the analysis. Note further that in the analysis we use the subnational growth rates directly without aggregating.

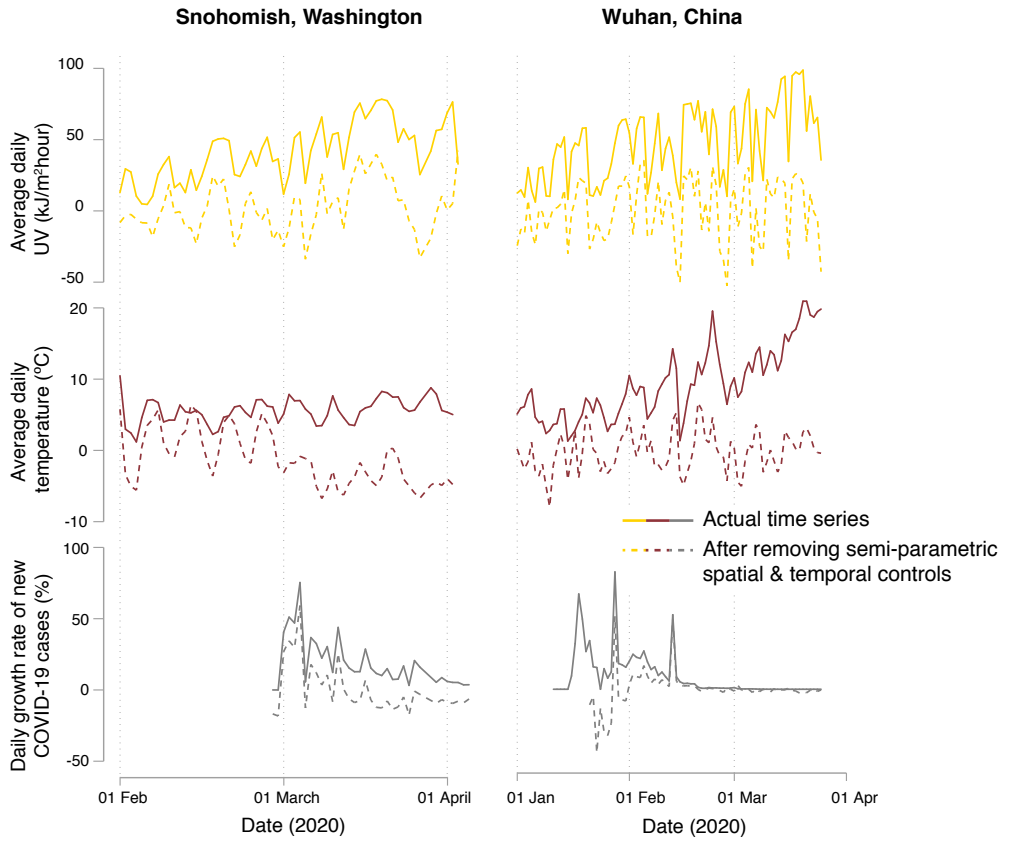

Figure S3: Semi-parametric controls remove potentially confounding seasonal variation. Our empirical strategy isolates idiosyncratic variation in local climatological conditions through the inclusion of semi-parametric controls in a panel regression framework (Eq. S1). This figure provide two illustrations of how general trends in climate conditions (e.g. gradual warming over the winter months) and in the evolution of COVID-19 are removed through the use of these controls. Solid lines indicate raw time series data from Snohomish, WA, United States (left) and from Wuhan, China (right) for UV exposure (gold), temperature (maroon), and daily COVID-19 growth rates (grey). Dotted lines show the residual variation used to causally estimate the effects of each weather variable; by removing location-specific averages, day-of-year averages, and country-specific weekly shocks, these time series no longer display trending behavior, thus removing the possibility that other, unobserved, trending variables may confound empirical estimates.

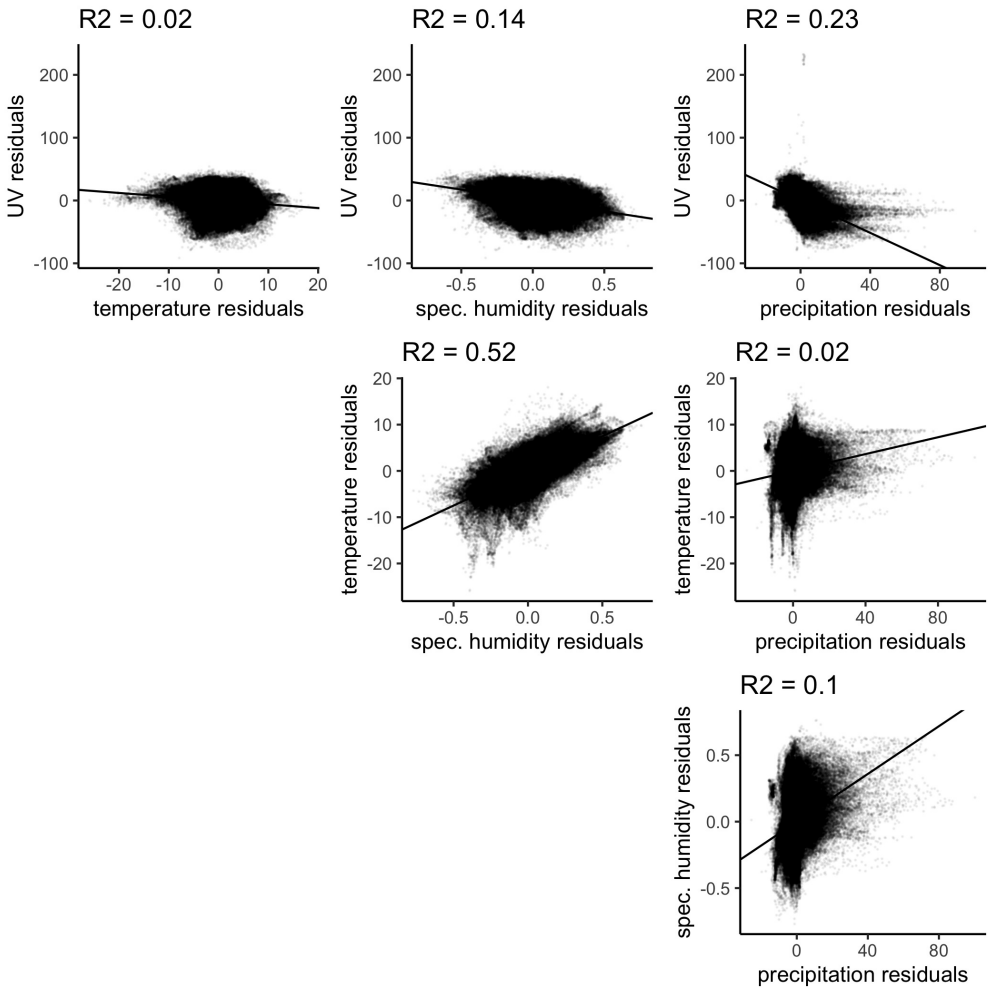

Figure S4: Correlation between daily environmental variables accounting for semi-parametric controls. Correlation between daily average UV (kJ/(m^2 hour)), average temperature (°C) , average humidity (%) and total precipitation (mm), after removing the semi-parametric controls in Eq. S1 (described in Section A.2). Linear fits are shown, with associated R^2 values.

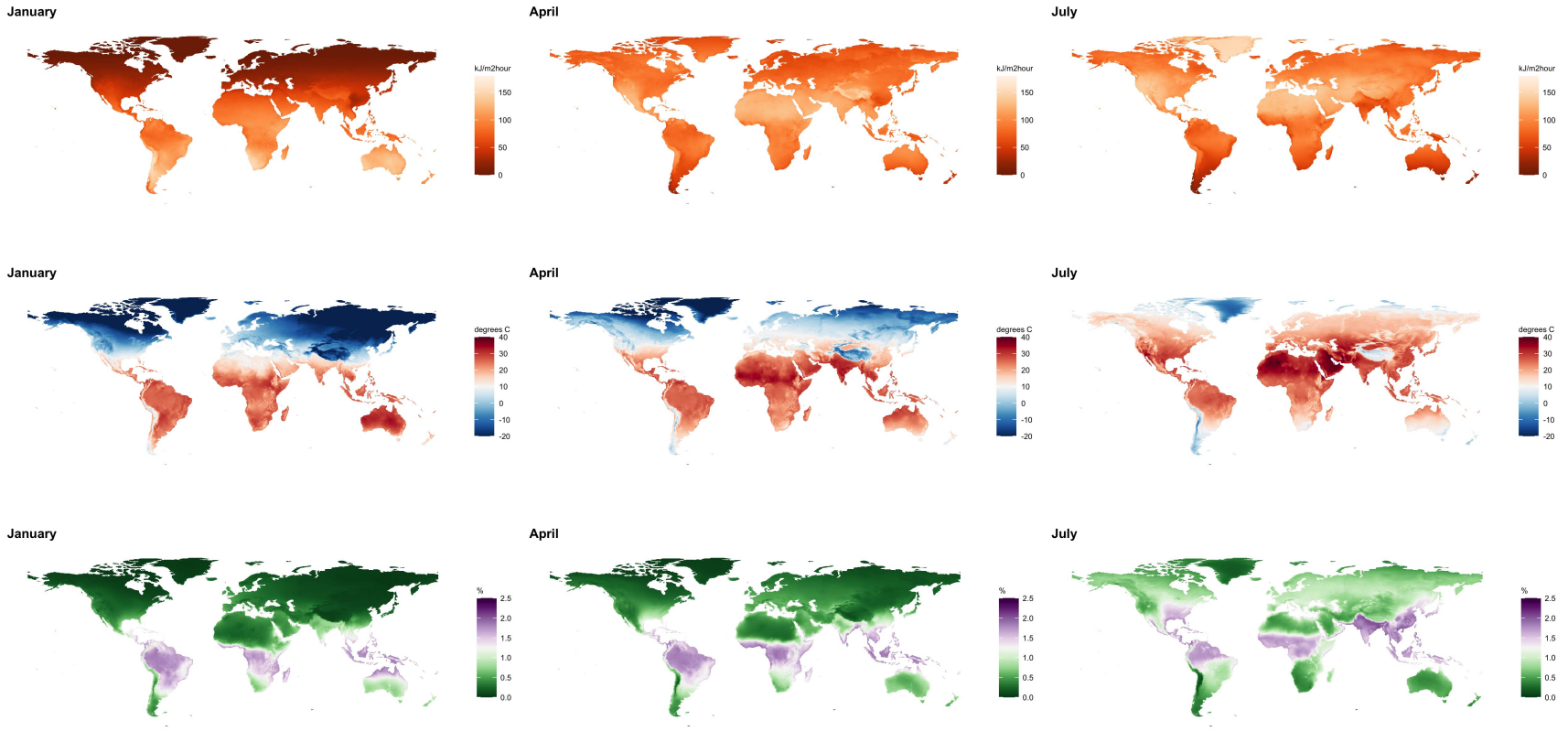

Figure S5: Seasonal changes in UV, temperature and specific humidity. Values show averages over 2015-2019 for UV (kJ/m^2 hour), temperature (° C) and specific humidity (%) (across rows) and across January, April and July (across columns) at a 0.25° latitude by 0.25° longitude resolution.

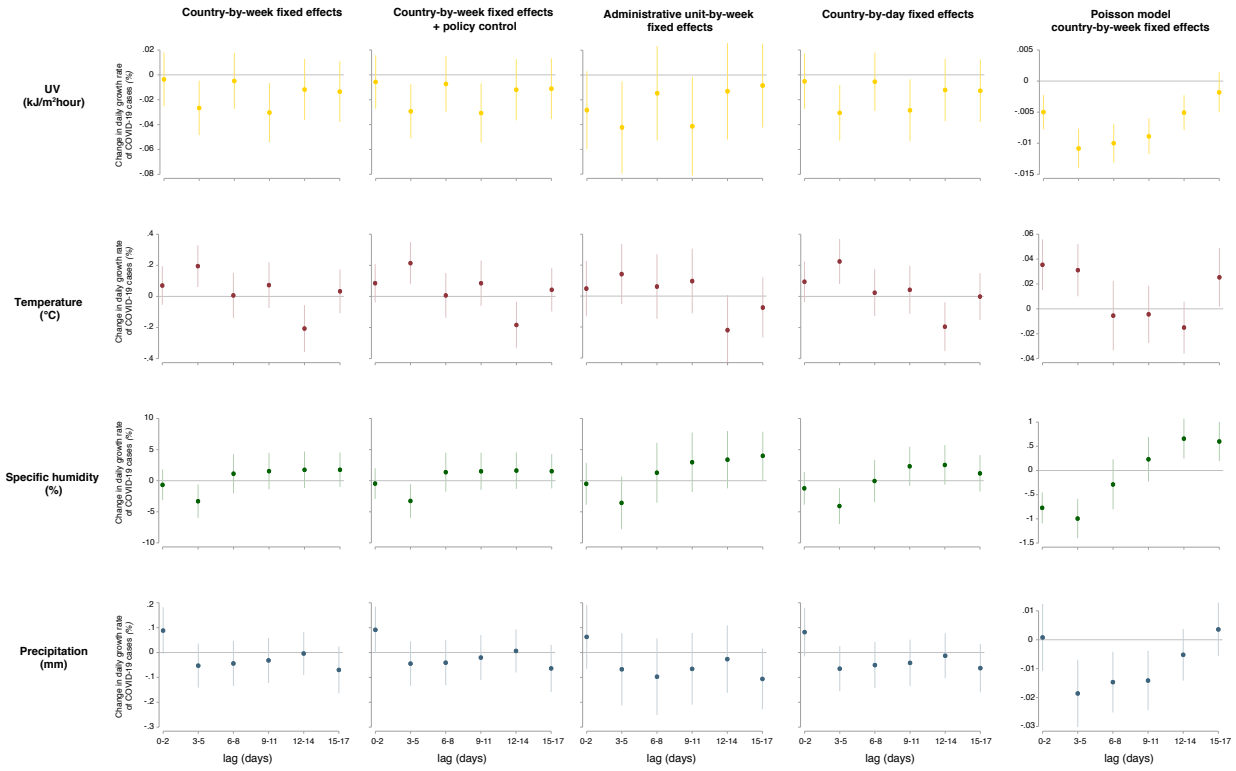

Figure S6: Alternative model specifications for the empirical relationship between daily growth rates of COVID-19 and climatological variables. Each column shows the estimated dynamic response of the daily growth rate in COVID-19 cases to UV (gold), temperature (maroon), specific humidity (green), and precipitation (blue) under a distinct set of semi-parametric controls. All columns include administrative unit (e.g. country, province, county) "fixed effects" (i.e. dummies), and day of year fixed effects and all responses in each column are estimated jointly in a single regression. Column 1 includes country by week of year fixed effects and is our baseline specification shown throughout the main text; column 2 includes country by week fixed effects and spatially and temporally-varying controls for social distancing policies (Section B.3); column 3 includes administrative unit (i.e. subnational units when available) by week fixed effects; column 4 drops all national-level data and adds county-by-day-of-year fixed effects. Finally, column 5 estimates a Poisson regression in which new cases per 1 million people are estimated as an exponential function of lagged climate variables, controlling for lagged total cumulative cases (Supplementary Materials Section A.2).

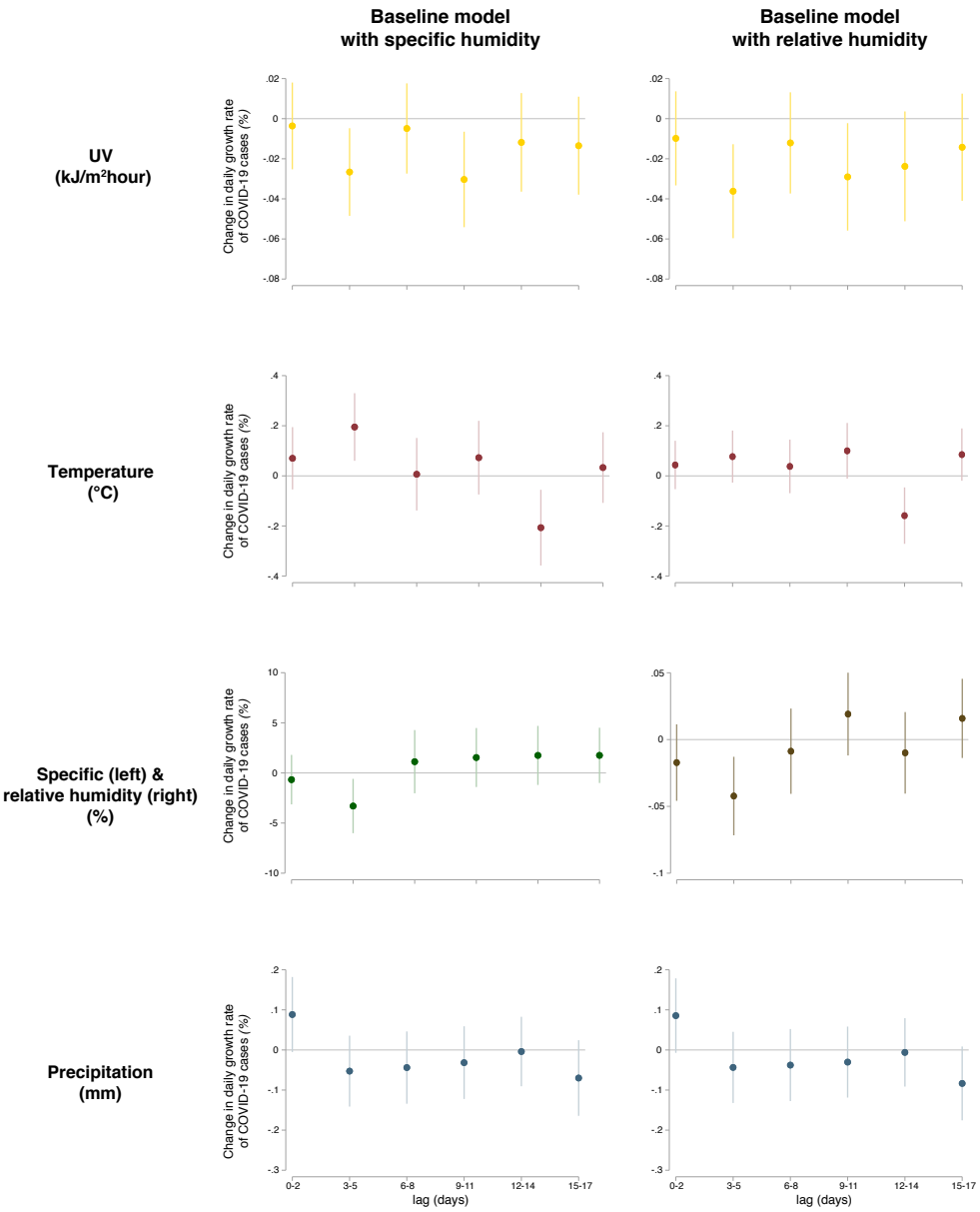

Figure S7: Empirical estimates of the dynamic relationship between COVID-19 and local climatic conditions using specific and relative humidity. Each column of this figure shows the estimated dynamic response of the daily growth rate in confirmed COVID-19 cases to lagged 3-day average UV (gold), temperature (maroon), specific humidity (green), relative humidity (brown), and precipitation (blue). The left column shows our baseline specification (Fig. 2A). The right column is identical, except that the model is estimated using relative humidity (%) instead of specific humidity (%). Note that these percentages represent conceptually different quantities. Relative humidity is the ratio of the partial pressure of water to the equilibrium vapor pressure, multiplied by 100, at a given temperature (i.e. what percent "full" of water is the air). Specific humidity gives the percent of an air parcel's total mass that is composed of water. The mean of relative humidity in our sample is 70% and the mean of specific humidity is 0.50%, which suggests that their estimated influences on the COVID-19 growth rate (i.e. the product of the estimated coefficients and changes in humidity) are of roughly similar magnitude.

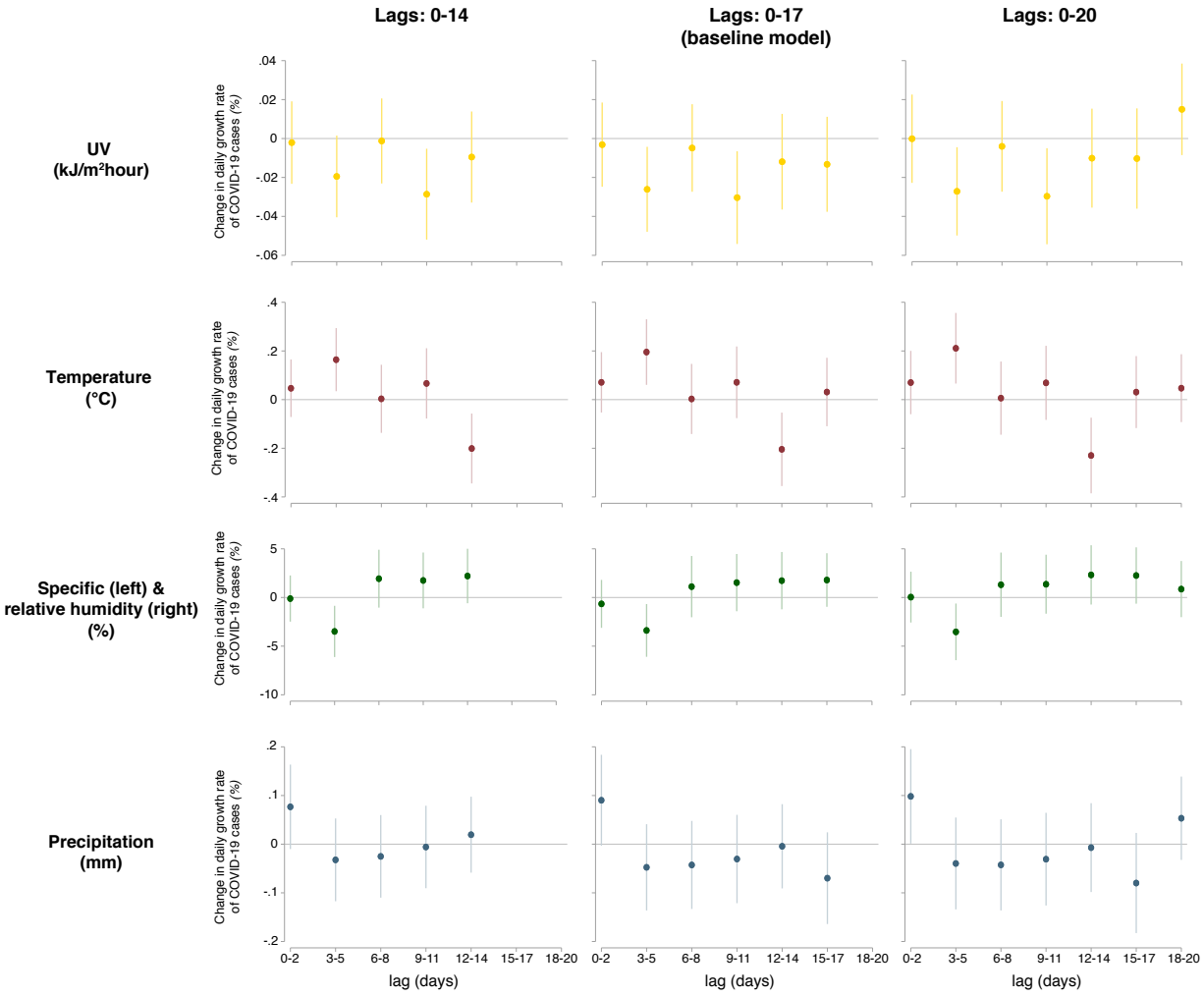

Figure S8: Empirical estimates of the dynamic relationship between COVID-19 and local climatic conditions using different distributed lag lengths. Each column of this figure shows the estimated dynamic response of the daily growth rate in confirmed COVID-19 cases to lagged 3-day average UV (gold), temperature (maroon), specific humidity (green), and precipitation (blue) occurring up to 20 days prior. All coefficients in each column were estimated jointly in a statistical model leveraging a rich set of semi-parametric controls to isolate idiosyncratic variation in each weather variable (Supplementary Materials Section A.2). Point estimates are indicated by circles and 95% confidence intervals are indicated by vertical lines. The first row omits the 15-17 day lag, which is included in our baseline specification (Fig. 2A), the second row replicates our baseline specification, and the third row adds an additional 18-20 day lag.

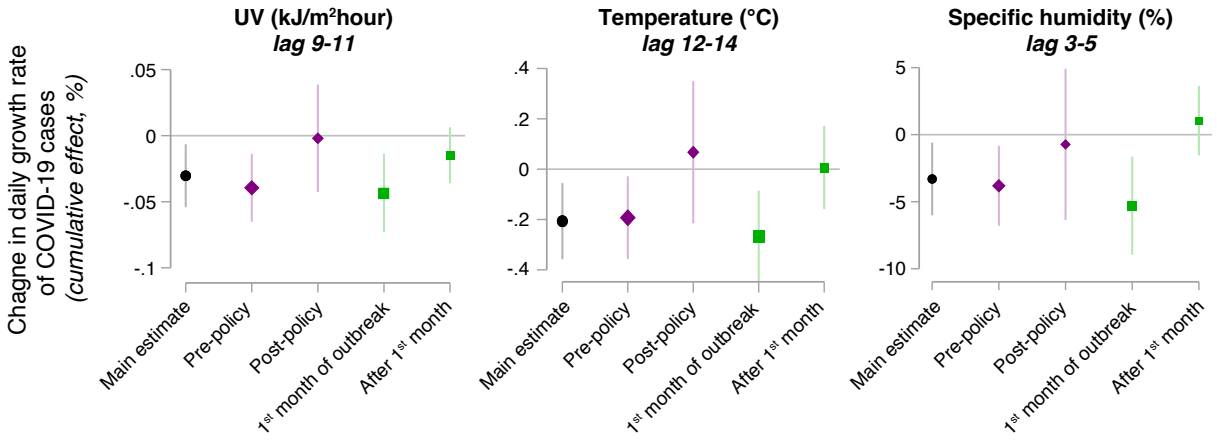

Figure S9: Individual lagged effects of climatological variables on daily growth rates of COVID-19 under distinct policy regimes and at different points in the outbreak. This figure shows the effect of select individual lag coefficients for each weather variable in black, with columns corresponding to UV (first column), temperature (second column), and specific humidity (third column). The lag shown for each climate variable is the largest magnitude effect recovered from a jointly estimated distributed lag regression including up to 17 days of lags. In purple, treatment effects of each weather variable are reported for the period of time before an administrative unit imposed any social distancing measures (large purple diamond), and after such measures were put in place (small purple diamond). Similarly, in green, treatment effects of each weather variable are reported for the first 30 days of the location-specific outbreak (large green square), and for all dates after the first 30 days (small green square). Vertical lines indicate 95% confidence intervals. Effects of social distancing policies and outbreak duration on the cumulative effect of all lagged coefficients for all three weather variables are shown in Fig. 2B.

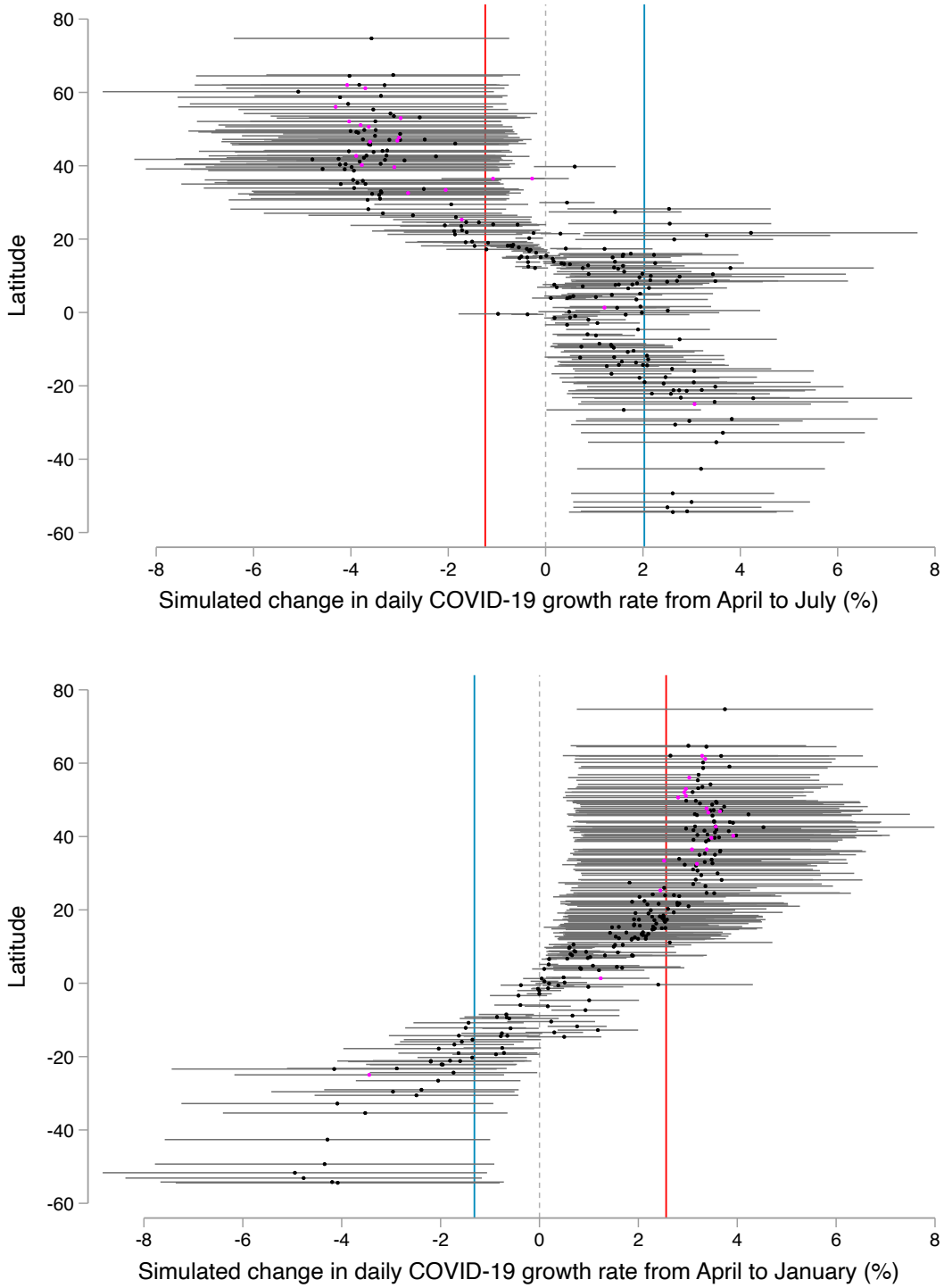

Figure S10: Simulated country-level daily COVID-19 growth rates by latitude. Top panel shows the projected percentage point change in the daily growth rate of COVID-2019 cases due to expected changes in UV between April 1, 2020 and July 31, 2020 at the country-level, where countries are ordered vertically by latitude based on area-weighted centroids. Circles show the simulated change in COVID-19 growth rate using our primary specification (Fig. 2). Red circles indicate 20 countries with highest COVID-19 cases per 1 million people, as of April 10, 2020. Grey error bars indicate 95% confidence intervals for each country. Bottom panel replicates top panel but for projected percentage point changes between April 1, 2020 and January 31, 2021. These projections reflect elevated or depressed growth rates driven only by expected UV changes.

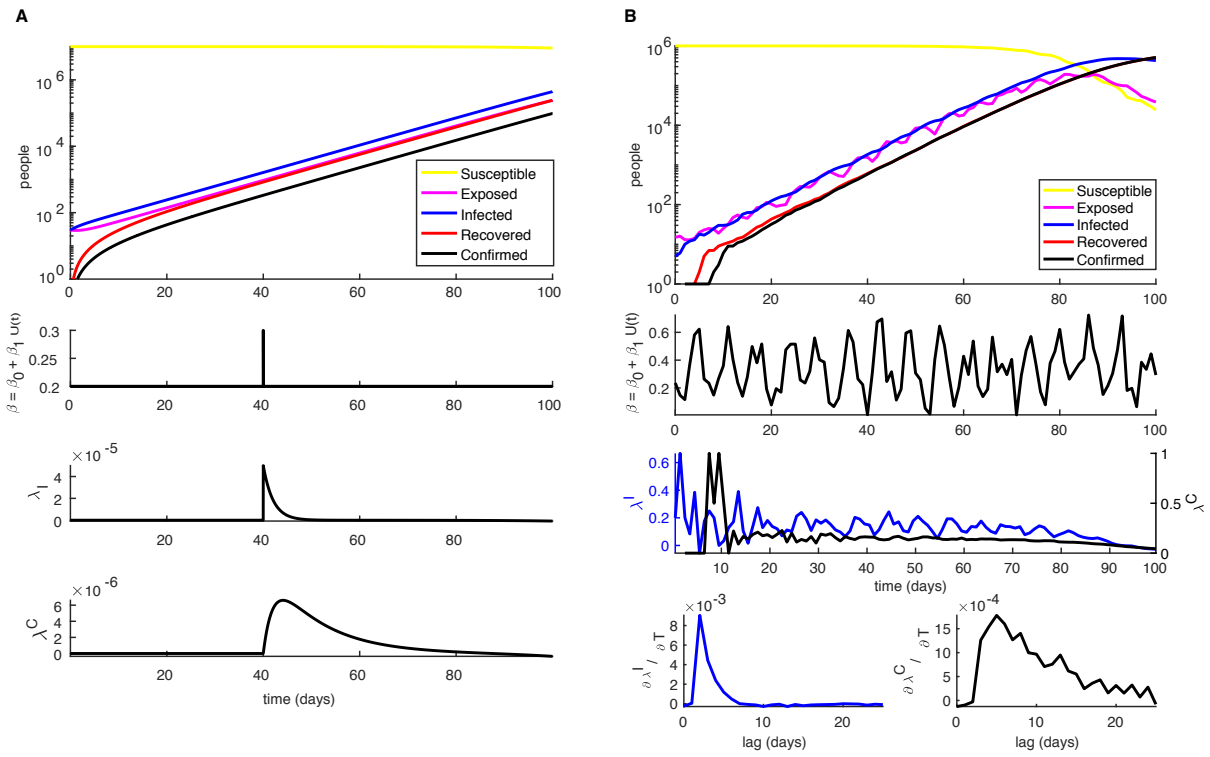

Figure S11: Simulated response of the growth rate of infectious and confirmed cases to time-varying transmission in the SEIR model. We let transmission vary over time with linear disturbances due to changes in weather U. In **A** we simulate the evolution of COVID-19 in the SEIR model deterministically using the semi-implicit Euler method. We parameterize U to generate a table-top perturbation, equal zero except for a single time step equaling one. In turn, this generates a temporary increase in β, which creates lagged increases in the growth rates of I and C, relative to a control run with constant β. In **B** we simulate the evolution of COVID-19 stochastically using the SEIR model to examine how a series of temperature-induced time-varying shocks to β impact λ^I and λ^C. We set U to be the sum of a sinusoid in t and Gaussian noise. Pooling observations from an ensemble of 300 runs, we estimate the effect of U on λ^I and λ^C using a distributed lag model and recover a lagged response of these growth rates to weather-induced changes in transmission. The goal of our empirical estimation is to test for such effects in real-world data.

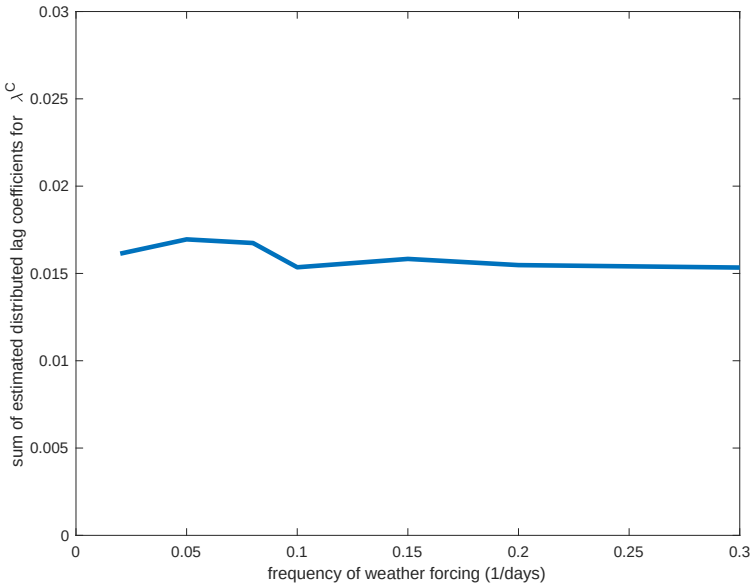

Figure S12: Estimated effect on case growth rate, λ_C, is similar for low and high frequency weather-induced shocks to transmission in stochastic SEIR simulations. We analyze the influence of the frequency of variation in U on the sensitivity of λ^I and λ^C to U using a stochastic SEIR model. We let transmission be a linear function of U (the linear effect of U on transmission is set to be 0.05) and parameterize U as sinusoidal a with constant amplitude, and variable frequency (from 0.02 to 0.3 $\frac{1}{days}$) across runs. Values in the plot show the sum of the coefficients of a distributed lag model of λ^C on U.

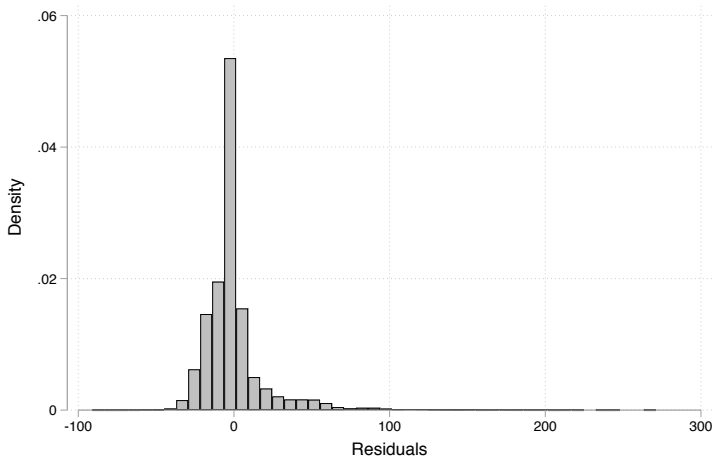

Figure S13: Distribution of residuals in daily growth rates. We estimate a distributed lag regression model in which the outcome variable is the growth rate in cumulative COVID-19 cases (Eq. S1). Here, we show the distribution of residuals after estimation of Eq. S1 using our baseline specification, which includes administrative unit fixed effects (i.e. dummy variables), country by week of year fixed effects, and day of year fixed effects (regression results shown in col. 3 of Table S1).

D Supplementary Tables

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) Poisson
UV kJ/m^{2}hour (lead: 1-3)				0.011				
				(0.012)				
UV kJ/m^{2}hour (lag: 0-2)	0.003	0.004	-0.004	-0.000	-0.006	-0.003	-0.028*	-0.005***
	(0.011)	(0.011)	(0.011)	(0.012)	(0.011)	(0.011)	(0.016)	(0.001)
UV kJ/m^{2}hour (lag: 3-5)	-0.025**	-0.017	-0.027**	-0.025**	-0.029***	-0.027**	-0.042**	-0.011***
	(0.012)	(0.011)	(0.011)	(0.012)	(0.011)	(0.011)	(0.019)	(0.002)
UV kJ/m^{2}hour (lag: 6-8)	-0.003	0.005	-0.005	-0.005	-0.007	-0.005	-0.015	-0.010***
	(0.012)	(0.011)	(0.011)	(0.013)	(0.011)	(0.011)	(0.019)	(0.002)
UV kJ/m^{2}hour (lag: 9-11)	-0.028**	-0.025**	-0.030**	-0.026*	-0.031**	-0.030**	-0.041**	-0.009***
	(0.013)	(0.012)	(0.012)	(0.014)	(0.012)	(0.012)	(0.020)	(0.001)
UV kJ/m^{2}hour (lag: 12-14)	-0.009	-0.011	-0.012	-0.003	-0.012	-0.011	-0.013	-0.005***
	(0.013)	(0.013)	(0.013)	(0.014)	(0.013)	(0.013)	(0.020)	(0.001)
UV kJ/m^{2}hour (lag: 15-17)	-0.009	-0.006	-0.013	-0.017	-0.011	-0.013	-0.009	-0.002
	(0.013)	(0.012)	(0.012)	(0.014)	(0.012)	(0.012)	(0.017)	(0.002)
Temp. °C (lead: 1-3)				0.043				
				(0.075)				
Temp. °C (lag: 0-2)	-0.016	0.030	0.070	0.055	0.085	0.069	0.048	0.035***
	(0.066)	(0.063)	(0.063)	(0.079)	(0.063)	(0.064)	(0.091)	(0.010)
Temp. °C (lag: 3-5)	0.119*	0.180***	0.195***	0.163**	0.214***	0.191***	0.143	0.031***
	(0.070)	(0.069)	(0.069)	(0.076)	(0.069)	(0.069)	(0.099)	(0.011)
Temp. °C (lag: 6-8)	-0.041	0.035	0.007	0.041	0.006	0.008	0.062	-0.005
	(0.074)	(0.072)	(0.074)	(0.083)	(0.073)	(0.074)	(0.106)	(0.014)
Temp. °C (lag: 9-11)	0.015	0.063	0.073	0.036	0.085	0.073	0.097	-0.004
	(0.078)	(0.074)	(0.075)	(0.084)	(0.075)	(0.075)	(0.107)	(0.012)
Temp. °C (lag: 12-14)	-0.189**	-0.188**	-0.207***	-0.172**	-0.184**	-0.207***	-0.220*	-0.015
	(0.078)	(0.075)	(0.077)	(0.083)	(0.076)	(0.077)	(0.116)	(0.011)
Temp. °C (lag: 15-17)	0.044	-0.002	0.033	0.056	0.043	0.031	-0.074	0.025**
	(0.071)	(0.070)	(0.072)	(0.076)	(0.071)	(0.072)	(0.099)	(0.012)
Spec. humd. % (lead: 1-3)				-1.353				
				(1.427)				
Spec. humd. % (lag: 0-2)	0.888	0.961	-0.668	0.826	-0.465	-0.756	-0.501	-0.772***
	(1.285)	(1.206)	(1.259)	(1.515)	(1.259)	(1.262)	(1.720)	(0.163)
Spec. humd. % (lag: 3-5)	-1.661	-3.417**	-3.302**	-2.488	-3.256**	-3.229**	-3.568*	-0.993***
	(1.417)	(1.358)	(1.382)	(1.606)	(1.379)	(1.386)	(2.158)	(0.208)
Spec. humd. % (lag: 6-8)	2.325	-0.690	1.124	1.544	1.367	1.043	1.284	-0.288
	(1.627)	(1.579)	(1.609)	(1.791)	(1.601)	(1.612)	(2.462)	(0.265)
Spec. humd. % (lag: 9-11)	2.160	-0.136	1.536	1.862	1.515	1.600	2.969	0.232
	(1.531)	(1.477)	(1.501)	(1.593)	(1.523)	(1.507)	(2.441)	(0.235)
Spec. humd. % (lag: 12-14)	0.734	0.520	1.749	1.822	1.634	1.713	3.378	0.658***
	(1.459)	(1.457)	(1.503)	(1.610)	(1.494)	(1.507)	(2.347)	(0.210)
Spec. humd. % (lag: 15-17)	0.158	1.609	1.756	0.531	1.526	1.773	4.006**	0.600***
	(1.373)	(1.378)	(1.407)	(1.532)	(1.410)	(1.411)	(1.974)	(0.206)
Observations	51139	51140	51126	43626	51126	50740	48328	45292
R-squared	0.210	0.218	0.223	0.249	0.224	0.223	0.332	
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Days since outbk. ctl.	Yes	No	No	No	No	No	No	No
Country trend	No	Yes	No	No	No	No	No	No
CntryXweek FE	No	No	Yes	Yes	Yes	Yes	No	Yes
Policy ctl.	No	No	No	No	Yes	No	No	No
Testing ctl.	No	No	No	No	No	Yes	No	No
AdminXweek FE	No	No	No	No	No	No	Yes	No
CntryXday FE	No	No	No	No	No	No	No	No
Pseudo R-squared								0.742

Table S1: Empirical estimation of the relationship between COVID-19 and climatological variables. Columns (1)-(8) show estimates of the distributed lag regression model from Eq. (S1) using daily longitudinal data across a pooled sample of national and subnational data (Fig. 1). The outcome is the daily growth rate of cumulative confirmed cases for columns (1) through (8). In column (9), a Poisson distributed lag regression model is used (Eq. S2). All models include administrative unit (e.g. country, province, or county) and day of year fixed effects, and all control for distributed lags in daily precipitation (in mm) and specific humidity (in %). Columns (1)-(8) include distinct semi-parametric and other controls: (1) a "fixed effect" (dummy variable) for the number of days since the outbreak began; (2) linear country-specific time trend; (3) country by week fixed effects; (4) country by week fixed effects, including leads of climate variables; (5) country by week fixed effects, including controls for temporally and spatially-varying social distancing policy controls (Section B.3); (6) country by week fixed effects, including a control for the stringency of COVID-19 testing at country level (Section B); (7) administrative unit (e.g. country, province, county) by week fixed effects; (8) country by day fixed effects. Standard errors clustered at the administrative unit level are in parentheses. P-values from two-sided t-tests with *** p<0.01, ** p<0.05, * p<0.1.

Country	Resolution	Units	Time coverage	Source	Variable obtained	Note
Austria	Bundesland	9	2/26 - 4/8	Ministry of Health[1]	Cumulative cases	
Belgium	Region	3	1/30 - 4/6	Public Health Institute[2]	New cases	
Brazil	State	27	2/25 - 4/5	Ministry of Health[3]	New cases	
Chile	Region	16	3/2 - 4/6	Ministry of Health[4]	New cases	Date corrected
China	City	339	1/10 - 3/25	Health Professionals network[5]	Cumulative cases	Imputed missing values
France	Region	13	1/24 - 3/25	Various[6]	New cases	
Germany	Land	16	2/24 - 4/7	Public Health Institute[7]	Cumulative cases	Date corrected; Missing values
Iran	Province	31	2/19 - 3/22	Ministry of Health, News Agency[8]	New cases	Imputed mising values
Italy	Province	107	2/24 - 4/7	Department of Civil Protection[9]	Cumulative cases	
Netherlands	Province	12	2/26 - 4/7	Public Health Institute [10]	New cases	Date corrected
Portugal	Region	7	3/2 - 4/7	Directorate-General of Health[11]	New cases	Date corrected
South Korea	Province	17	1/19 - 4/9	Public Health Institute[12]	New cases	Date corrected
Spain	A. C.	16	2/26 - 4/5	Ministry of Health[13]	Cumulative cases	Date corrected
Sweden	County	21	2/25 - 4/8	Public Health Agency[14]	New cases	Date corrected
United Kingdom	Country	4	1/30 - 4/5	Public Health Institute[15]	New cases	Date corrected
United States	County	2,438	1/21 - 4/5	Various[16]	Cumulative cases	

Table S2: COVID-19 cases data for subnational units. This table contains information about the COVID-19 cases database. It presents the time and spatial coverage of our subnational data, the sources and the variables provided by these sources. It also indicates the corrections we applied to the data.

[1] *Bundesministerium für Soziales, Gesundheit, Pflege und Konsumentschutz*, https://www.sozialministerium.at/Informationen-zum-Coronavirus/Neuartiges-Coronavirus-(2019-nCov).html; scraped from the Wikipedia article "COVID-19-Pandemie in Österreich" ("COVID-19 pandemic in Austria"), available at https://de.wikipedia.org/wiki/COVID-19-Pandemie_in_%C3%96sterreich.

[2] *Sciensano*, https://epistat.wiv-isp.be/covid/, obtained from the Wikipedia article "2020 coronavirus pandemic in Belgium", https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_Belgium

[3] *Ministério da Saúde*, https://covid.saude.gov.br/, obtained from the GitHub repository of Henrico Moraes, https://github.com/elhenrico/covid19-Brazil-timeseries

[4] *Ministerio de Salud*, https://www.minsal.cl/nuevo-coronavirus-2019-ncov/casos-confirmados-en-chile-covid-19/; scraped from the Wikipedia article "2020 coronavirus pandemic in Chile" (https://en.wikipedia.org/w/index.php?title=2020_coronavirus_pandemic_in_Chile).

[5] Dīng Xiāng Yuán (DXY), https://ncov.dxy.cn/ncovh5/view/en_pneumonia?from=dxy&source=&link=&share=; obtained from the Github repository of the Global Policy Lab: https://github.com/bolliger32/gpl-covid

[6] The compiled data have been scraped from the Wikipedia article"2020 coronavirus pandemic in France", https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_France.

[7] Robert Koch Institute, https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Fallzahlen.html. The data have been scraped from the Wikipedia article "2020 coronavirus pandemic in Germany", available at https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_Germany.

[8] https://en.irna.ir/photo/83723991/Iran-s-coronavirus-toll-update-March-22-2020. The compiled data have been scraped from the Wikipedia article "2020 coronavirus pandemic in Iran", available at https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_Iran.

[9] *Dipartimento della Protezione Civile*, http://opendatadpc.maps.arcgis.com/apps/opsdashboard/index.html#/b0c68bce2cce478eaac82fe38d4138b1. The data are directly available on their GitHub repository, at https://github.com/pcm-dpc/COVID-19.

[10] Institute for Public Health and the Environment, https://www.rivm.nl/en/novel-coronavirus-covid-19/current-information-about-novel-coronavirus-covid-19. The data are available on the Wikipedia page "2019-20 coronavirus pandemic data, Netherlands medical cases", https://en.wikipedia.org/wiki/Template:2019%E2%80%9320_coronavirus_pandemic_data/Netherlands_medical_cases.

[11] *Direção-Geral da Saúde*, https://covid19.min-saude.pt/relatorio-de-situacao/. The data have been scraped from the Wikipedia article "2020 coronavirus pandemic in Portugal" (https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_Portugal).

[12] Korean Center for Disease Control and Prevention, https://www.cdc.go.kr/board/board.es?mid=a30402000000&bid=0030. The data have been obtained from the Wikipedia article "2020 coronavirus pandemic in South Korea", available at https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_South_Korea.

[13] *Ministerio de Sanidad – Centro de Coordinación de Alertas y Emergencias Sanitarias*, https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov-China/situacionActual.htm. The data have been compiled by Datadista and made available on their GitHub repository, at https://github.com/datadista/datasets/tree/master/COVID%2019.

[14] *Folkhälsomyndigheten*, https://experience.arcgis.com/experience/09f821667ce64bf7be6f9f87457ed9aa. The data have been compiled from the Wikipedia article "2020 coronavirus pandemic in Sweden", (https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_Sweden).

[15] Public Health England, https://www.arcgis.com/apps/opsdashboard/index.html#/f94c3c90da5b4e9f9a0b19484dd4bb14. The data have been scraped from the Wikipedia article "2020 coronavirus pandemic in the United Kingdom", available at https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_the_United_Kingdom.

[16] All data have been meticulously compiled by the *New York Times*, and can be downloaded from the GitHub repository they dedicated to this issue: https://github.com/nytimes/covid-19-data/archive/master.zip.

Country	Resolution	Year	Source
Austria	Bundesland	January 2015	Statistik Austria[1]
Belgium	Region	January 2015	STATBEL[2]
Brazil	State	July 2019	Brazilian Institute of Geography and Statistics (IBGE)[3]
Chile	Region	2017	Instituto Nacional de Estadísticas (INE)[4]
China	City	2010	National Bureau of Statistics of China[5]
France	Region	January 2017	INSEE[6]
Germany	Land	December 2018	Statistisches Bundesamt[7]
Iran	Province	2016	Statistical Center of Iran[8]
Italy	Province	January 2019	ISTAT[9]
Netherlands	Province	November 2019	Statistics Netherlands (CBS)[10]
Portugal	Region	December 2018	Statistics Portugal (INE)[11]
South Korea	Province	December 2017	citypopulation.de[12]
Spain	Autonomous Community	2019	Spanish Statistical Office (INE)[13]
Sweden	County	December 2018	Statistics Sweden (SCB).[14]
United Kingdom	Country	2018	Office for National Statistics[15]
United States	County	July 2019	US Census Bureau[16]

Table S3: Population data at the subnational level. Our population data for each subnational unit come mostly from national offices of statistics, listed in this table. This table also indicates when the population estimates were made.

[1]Available at http://www.statistik.at/web_de/statistiken/menschen_und_gesellschaft/bevoelkerung/080912.html (in German). The data have been scraped from the Wikipedia article "States of Austria", available at https://en.wikipedia.org/w/index.php?title=States_of_Austria&oldid=947664334.

[2]Available at https://statbel.fgov.be/fr/themes/population/structure-de-la-population (in French).

[3]Available here: https://www.ibge.gov.br/en/statistics/social/population/18448-estimates-of-resident-population-for-municipalities-and-federation-units.html?=&t=resultados.

[4]Downloadable here: https://www.ine.cl/estadisticas/sociales/censos-de-poblacion-y-vivienda/poblacion-y-vivienda (in Spanish).

[5]http://www.stats.gov.cn/tjsj/ndsj/2010/indexeh.htm. The data have been downloaded from the GitHub repository of the Global Policy Lab, available at https://github.com/bolliger32/gpl-covid.

[6]Available here: https://www.insee.fr/fr/statistiques/4265390?sommaire=4265511 (in French). The data have been scraped from the Wikipedia article "Région française" and are available at https://fr.wikipedia.org/wiki/R%C3%A9gion_fran%C3%A7aise.

[7]Retrievable here: https://www-genesis.destatis.de/genesis/online?operation=abruftabelleBearbeiten&levelindex=2&levelid=1587183994407&auswahloperation=abruftabelleAuspraegungAuswaehlen&auswahlverzeichnis=ordnungsstruktur&auswahlziel=werteabruf&code=12411-0010&auswahltext=&werteabruf=Value+retrieval#astructure. The data have been scraped from the Wikipedia article "List of German States by Population" (https://en.wikipedia.org/wiki/List_of_German_states_by_population).

[8]https://www.amar.org.ir/english/Population-and-Housing-Censuses. The data have been scraped from the Wikipedia article "Provinces of Iran", available at https://en.wikipedia.org/wiki/Provinces_of_Iran.

[9]Available at http://demo.istat.it/pop2019/index3_e.html. The data have been downloaded from the GitHub repository of the Global Policy Lab, available at https://github.com/bolliger32/gpl-covid.

[10]Available here: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37230ned/table (in Dutch). The data have been scraped from the Wikipedia article "Provinces of the Netherlands", available at https://en.wikipedia.org/wiki/Provinces_of_the_Netherlands.

[11]Downloadable here: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_publicacoes&PUBLICACOESpub_boui=411651997&PUBLICACOESstema=00&PUBLICACOESmodo=2.

[12]https://www.citypopulation.de/en/southkorea/cities/

[13]Available at https://www.ine.es/jaxiT3/Tabla.htm?t=4925&L=0 (in Spanish)

[14]The original data can be found at https://www.scb.se/en/finding-statistics/statistics-by-subject-area/population/population-composition/population-statistics/pong/tables-and-graphs/quarterly-population-statistics--municipalities-counties-and-the-whole-country/julydecember-2018/.

[15]Complete report available here: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2018.

[16]Downloadable at: https://www2.census.gov/programs-surveys/popest/tables/2010-2019/counties/totals/co-est2019-annres.xlsx.