Climate Affects Global Patterns Of Covid-19 Early Outbreak Dynamics

Gentile Francesco Ficetola[1,2,*] & Diego Rubolini[1*]

1 Dipartimento di Scienze e Politiche Ambientali, Università degli Studi di Milano, via Celoria 26, I- 20133 Milano, Italy

2 Université Grenoble Alpes, CNRS, Université Savoie Mont Blanc, LECA, Laboratoire d'Ecologie Alpine, F-38000 Grenoble

*Contributed equally to this work; order was decided with a coin toss

Correspondence: francesco.ficetola@gmail.com, diego.rubolini@unimi.it

Phone numbers
Gentile Francesco Ficetola: 0039 338 1844255
Diego Rubolini: 0039 02 503 14718

Abstract

Environmental factors, including seasonal climatic variability, can strongly impact on spatio-temporal patterns of infectious disease outbreaks. We assessed the effects of temperature and humidity on the global patterns of COVID-19 early outbreak dynamics during January-March 2020. Climatic variables were the best drivers of global variation of confirmed COVID-19 cases growth rates. Growth rates peaked in temperate regions of the Northern Hemisphere with mean temperature of ~5°C and humidity of ~0.6-1.0 kPa during the outbreak month, while they decreased in warmer and colder regions. The strong relationship between local climate and COVID-19 growth rates suggests the possibility of seasonal variation in the spatial pattern of outbreaks, with temperate regions of the Southern Hemisphere becoming at particular risk of severe outbreaks during the next months.

A recently discovered coronavirus, SARS-CoV-2, is the aethiological agent of a pandemic disease, Covid-19, causing severe pneumonia outbreaks at the global scale[1]. Covid-19 cases are now reported in more than 155 countries and regions worldwide[2]. Three months after the discovery of SARS-CoV-2, the global pattern and the early dynamics of Covid-19 outbreaks seem highly variable. Some countries have been experiencing limited growth and spread of Covid-19 cases, while others are suffering widespread community transmission and nearly exponential growth of infections[2]. Given the impact of environmental conditions on the transmission of many pathogens, we tested the hypothesis that the severity of Covid-19 outbreaks across the globe is affected by spatial variation of key environmental factors, such as temperature, air humidity and pollution[3-7]. We then evaluated if this could help to illustrate global variation in the risk of severe Covid-19 outbreaks in the coming months.

Relying on a publicly available global dataset[2], we computed the daily growth rates *r* of confirmed Covid-19 cases (Covid-19 growth rate hereafter) for 121 countries/regions (see the Methods section and Table S6 in the Supplementary Appendix). We limited our measure of epidemics growth rate to the first 5 days after reaching a minimum threshold of confirmed cases (25, 50 or 100), as the mean incubation period of Covid-19 is ca. 5 days[8] and, immediately after the first confirmed cases, many countries put in place unprecedented containment measures to mitigate pathogen spread and community transmission[9]. Variation at these early epidemic growth rates should best reflect the impact of local environmental conditions on disease spread. We restricted analyses to data reported before March 19, as during that week many regions of the world adopted stringent containment measures even in absence of large numbers of reported cases. For instance, on March 17, 37 US states closed schools to prevent disease spread, including several states with less than 25 confirmed Covid-19 cases[10]. We also considered additional factors that could affect SARS-CoV-2 transmission dynamics, such as human population density, government per-capita health expenditure, and average air pollution levels (fine particulate matter; see Methods).

Covid-19 growth rates showed high variability at the global scale (Fig. 1A-C). The observed daily growth rate after reaching 50 cases (r_{50}) was on average 0.18 [95% CI 0.16-0.19], and ranged from 0.01 (Kuwait) to 0.44 (Denmark). The highest growth rates were observed in temperate regions of the Northern Hemisphere (Fig. 1C). Growth rates calculated using different minimum thresholds of confirmed cases (25 or 100) were strongly positively correlated (see Methods), indicating robustness of our results to the choice of thresholds.

Climate variables were the most important predictors of Covid-19 growth rate (Table S1). The best-fitting linear mixed model suggested that r_{50} is non-linearly related to spatial variation in mean temperature of the outbreak month (Fig. 1A, Tables S2-S3). Growth rates peaked in regions with mean temperature of ~5°C during the outbreak month, and decreased both in warmer and colder climates (Fig. A, Table S3). The comparison of models with different combinations of predictors confirmed temperature as the variable with the highest relative importance in explaining variation of r_{50} (Table S1), and temperature was the only parameter included in the best-fitting model (Tables S2-S3). Temperature and humidity of the outbreak month showed a strong, positive relationship across regions (Fig. S1), thus they could not be included as predictors in the same model. When we repeated the analyses including humidity instead of temperature, r_{50} varied significantly and non-linearly with humidity, peaking at ~0.6-1.0 kPa (Fig. 1B, Tables S4-S5). The best model also showed slightly larger growth rates in countries with greater health expenditure (Table S5), possibly because of more efficient early reporting and/or faster diagnosis of Covid-19 cases.

Results were highly consistent if we calculated growth rates after minimum thresholds of 25 or 100 cases (r_{25} and r_{100}, respectively) instead of 50 (Tables S3 and S5). Human population density and air pollution showed very limited relative importance values (always < 0.50; Table S1).

We then displayed potential seasonal changes in Covid-19 growth rates by projecting our best model of r_{50} in relation to temperature under the average temperature conditions of the current (March) and next (June and September) months (Fig. 1C-E). The predicted global distribution of Covid-19 growth rates based on March temperatures showed favorable conditions for disease spread in most temperate regions of the Northern Hemisphere, and matched well with the observed spatial distribution of Covid-19 growth rates during the January-March global outbreak (Fig. 1C). The expected seasonal rise in temperatures during the next months could results in less suitable conditions for Covid-19 spread in these areas. Conversely, seasonally decreasing temperatures could accelerate disease spread in large areas of the Southern Hemisphere, including south America, south Africa, eastern Australasia and New Zealand (Fig. 1D-E).

The management of Covid-19 outbreaks is undoubtedly one of the biggest challenges governments will face in the coming months. Our spatially-explicit analysis suggests that, at least in some parts of the world, ongoing containment efforts could benefit from the interplay between pathogen spread and local climate. We do not claim that climate is the single major driver of Covid-19 spread. The huge variation of Covid-19 growth rates among regions with similar climate indeed suggests that diverse and complex social and demographic factors, as well as stochasticity, may strongly contribute to determine the severity of Covid-19 outbreaks. Yet, climate can contribute to explain the variability in global patterns of Covid-19 growth rates. In the coming months, we may thus expect that large areas of the Southern Hemisphere will show environmental conditions promoting severe Covid-19 outbreaks.

Materials and methods

Covid-19 dataset

We downloaded the time series of confirmed Covid-19 cases from the Johns Hopkins University Center For Systems Science and Engineering (JHU-CSSE) GitHub repository (https://github.com/CSSEGISandData/Covid-19/; file 'time_series_19-covid-Confirmed.csv') [11]. This datafile is updated once a day (at 23:59 UTC) and reports, for each day since January 22, 2020, all confirmed Covid-19 cases at the country level or at the level of significant geographical units belonging to the same country, which we defined here as 'regions' (e.g. US states or China provinces), whenever separate Covid-19 cases data for these regions are available. Initially, US data were reported by county but, as of March 9, they were reported at the state level. We therefore merged all US county data before March 9 to state level, and used state-level time series for subsequent calculations. With the exception of US data, in all other cases we maintained the original country/region information adopted by the JHU-CSSE. The datafile considered for the analyses was downloaded on March 19, 2020, and included confirmed Covid-19 cases until March 18, 2020. From this dataset, we selected data for all countries / regions for which local outbreaks were detected. We define a local outbreak event when at least 50 positive cases were detected in a given country / region, and calculated the growth rate of confirmed Covid-19 cases between day 1 and day 5, when day 1 was the day at which the 50 cases threshold was reached. We calculate the daily growth rate r of confirmed Covid-19 cases for each country/region, assuming an exponential growth as: $r = [\ln(\text{n cases}_{\text{day 5}}) - \ln(\text{n cases}_{\text{day 1}})] / 5$. We checked the robustness of our estimates of growth rate by calculating daily growth rate after the first 25, 50 or 100 cases (r_{25}, r_{50} and r_{100}, respectively). Growth rates estimated at different thresholds were strongly positively correlated (Pearson's correlation coefficients, r_{25} vs. r_{50}: r = 0.74; r_{50} vs. r_{100}: r = 0.81).

The dataset does not report information on containment measures, and these may be highly heterogeneous among countries/regions. We decided to calculate growth rate on the basis of the first five days, in order to obtain an estimate of the non-intervened spread of the disease (i.e. before stringent containment measures are undertaken). Five days provides a reasonable trade-off between having to unreliable estimates of growth rates (if calculated on the basis of a smaller number of days, e.g. 3), and obtaining growth rates influenced by the enforcement of heavy containment measures (such as immediate isolation of confirmed cases). Five days is the median estimated time spanning before the onset of symptoms [8], implying that infected patients might spread the virus for 5 days undetected in absence of preventive control measures. The mean estimated growth rate of confirmed Covid-19 cases showed a tendency to decrease from r_{25} to r_{100} (mean and 95% c.i.: r_{25} = 0.21 [0.19-0.22, n = 121], r_{50} = 0.18 [0.16-0.19, n = 90], r_{100} = 0.16 [0.14-0.18, n = 69]), possibly because of the progressive effect of containment measures that were adopted in different countries at different times and different minimum thresholds after the onset of the local outbreak. We excluded from analyses countries/regions with less than 100000 inhabitants (in our dataset, San Marino only). As of March 19, 2020, the JHU-CSSE dataset provided information for a total of 121 countries/regions for the calculation of r_{25}, 90 for r_{50}, and 69 for r_{100}. The final list of countries/regions included in the analyses, together with estimated confirmed Covid-19 growth rates at different thresholds, is reported in Table S6.

Environmental and socio-economic variables

We considered two climatic variables that are known to affect the spread of viruses: mean air temperature and vapor pressure, which is a measure of absolute humidity. Previous studies showed that, for coronaviruses and influenza viruses, survival is generally higher at low temperature and low values of absolute humidity [5,6,12-14]. For each country/region, we thus calculated the mean monthly values for temperature (°C) and vapor pressure (kPa) for January, February and March on the basis of the WorldClim 2.1 raster layers at 10 arc-minutes resolution [15]. We relied on WorldClim climatic data because homogeneous data on conditions for the period January-March 2020 are not yet available at a global scale (see e.g.

https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5), and spatial variation among areas of the world is generally much stronger than inter-annual variation for the same region [16]. As additional predictors, we considered mean human population density [17] (population density hereafter, expressed in inhabitants/km^2) and per-capita government health expenditure (health expenditure hereafter) (indicator 'Domestic General Government Health Expenditure (GGHE-D) per Capita in US$; average of 2015-2017 values downloaded from the World Health Organization database at https://apps.who.int/nha/database). Health expenditure was available at country-level only: hence, regions within countries were assigned the same health expenditure value. Finally, it has been proposed that air pollution, and especially fine atmospheric particulate, could enhance the persistence and transmission of coronaviruses [3,18]. We therefore extracted values of annual concentration (μg/m^3) of ground-level fine particulate matter (PM2.5) for 2016 from the NASA Socioeconomic Data and Applications Center [19], and calculated the mean abundance of PM2.5 for each country/region. We performed all spatial analyses using the raster package in R [20].

Statistical analyses

We used linear mixed models (LMMs) to relate the global variation of r_{50}, r_{25} and r_{100} to the five environmental predictors (temperature and humidity of outbreak month; population density; health expenditure and PM2.5). To associate climate variables to the estimated *r*-values for each country/region, we first extracted the mean month of the 5 days over which we computed the *r*-values (rounded to the nearest integer) (outbreak month). We then assigned to the *r*-values of each country/region the mean temperature and humidity of the month during which the outbreak occurred. Country was included as a random factor to take into account potential non-independence of growth rates from regions belonging to the same country. Non-linear relationships between climatic factors and ecological variables are frequent [21], and in exploratory plots we detected a clear non-linear relationship between *r*-values and climate. Therefore, for climatic variables, we included in models both linear and quadratic terms. Humidity, population density, health expenditure and PM2.5 were $\log_{10}$-transformed to reduce skewness and improve normality of model residuals.

We adopted a model selection approach to identify the variables most likely to affect the global variation of Covid-19 growth rate [22]. We built models representing the different combinations of independent variables, and ranked them on the basis of Akaike's Information Criterion (AIC). AIC trades-off explanatory power vs. number of predictors; parsimonious models explaining more variation have the lowest AIC values and are considered to be the "best models" [22]. For each candidate model, we calculated the Akaike weight ω_i, representing the probability of the model given the data [23]. We then calculated the relative variable importance of each variable (RVI) as the sum of ω_i of the models where each variable is included. RVI can be interpreted the probability that a variable should be included in the best model [22,24]. Model selection analyses and the calculation of RVI can be heavily affected by collinearity among variables. In our dataset, temperature and humidity showed a very strong positive correlation (Fig. S1 and Table S7); furthermore, population density was strongly positively associated with PM2.5 (Figure S1 and Table S7). Therefore, temperature and humidity, or population density and PM2.5, could not be considered together in the same models [24,25]. All other predictors showed weak correlations and should not cause collinearity issues [25] (Table S7). We therefore repeated the model selection for different combinations of uncorrelated variables. First, we considered temperature, health expenditure and population density as independent variables. Then we repeated the analysis using humidity instead of temperature, and we calculated the RVI of variables separately for these two model selection analyses. Finally, to assess the role of PM2.5, we repeated these two model selections analyses using PM2.5 instead of health expenditure. The RVI values for all tested models are reported in Table S1. Due to low RVI of PM2.5 in all models (Table S1), we subsequently report detailed results of models including population density instead of PM2.5 (Tables S2-S5). To test the robustness of our conclusion to subjective thresholds for the minimum number of cases, all analyses were repeated considering the three estimates of Covid-19 growth rate as dependent variables (r_{25}, r_{50} and r_{100}).

LMMs were fitted using the lmer function of the lme4 R package [26], while tests statistics were calculated using the lmerTest package [27]. To confirm that spatial autocorrelation did not bias the outcome of our analyses, we calculated the spatial autocorrelation (Moran's *I*) of the residuals of best-fitting models using the EcoGenetics package in R [28] at lags of 1000 km up to a maximum distance of 5000 km. Model residuals did not show significant spatial autocorrelation at any lag (in all cases, |Moran's I|

< 0.10 and P > 0.11), suggesting that spatial autocorrelation was not a major issue in our analyses [29].

References

1. World Health Organization. Coronavirus disease (COVID-2019) situation reports: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/ [accessed 1 March 2020]; 2020.

2. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases 2020:DOI:https://doi.org/10.1016/S473-3099(20)30120-1. Data available at https://github.com/CSSEGISandData/COVID-19.

3. Setti L, Passarini F, de Gennaro G, et al. Relazione circa l'effetto dell'inquinamento da particolato atmosferico e la diffusione di virus nella popolazione. http://www.simaonlus.it/wpsima/wp-content/uploads/2020/03/COVID19_Position-Paper_Relazione-circa-l%E2%80%99effetto-dell%E2%80%99inquinamento-da-particolato-atmosferico-e-la-diffusione-di-virus-nella-popolazione.pdf: Società Italiana di Medicina Ambientale; 2020.

4. Sajadi MM, Habibzadeh P, Vintzileos A, Shokouhi S, Miralles-Wilhelm F, Amoroso A. Temperature and Latitude Analysis to Predict Potential Spread and Seasonality for COVID-19. 2020:Available at SSRN: https://ssrn.com/abstract=3550308.

5. Kampf G, Todt D, Pfaender S, Steinmann E. Persistence of coronaviruses on inanimate surfaces and their inactivation with biocidal agents. Journal of Hospital Infection 2020;104:246-51.

6. Tamerius JD, Shaman J, Alonso WJ, et al. Environmental Predictors of Seasonal Influenza Epidemics across Temperate and Tropical Climates. PLoS Pathog 2013;9:12.

7. Wang J, Tang K, Feng K, Lv W. High Temperature and High Humidity Reduce the Transmission of COVID-19 (March 9, 2020). Available at SSRN 2020:http://dx.doi.org/10.2139/ssrn.3551767

8. Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. New England Journal of Medicine 2020.

9. Hellewell J, Abbott S, Gimma A, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. The Lancet Global Health 2020;8:e488-e96.

10. Maxouris C. These states have some of the most drastic restrictions to combat the spread of coronavirus. CNN 2020:https://edition.cnn.com/2020/03/17/us/states-measures-coronavirus-spread/index.html.

11. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases 2020.

12. Shaman J, Kohn M. Absolute humidity modulates influenza survival, transmission, and seasonality. Proceedings of the National Academy of Sciences 2009;106:3243.

13. Lowen AC, Steel J. Roles of Humidity and Temperature in Shaping Influenza Seasonality. Journal of Virology 2014;88:7692-5.

14. Lowen AC, Mubareka S, Steel J, Palese P. Influenza virus transmission is dependent on relative humidity and temperature. PLoS Pathog 2007;3:1470-6.

15. Fick SE, Hijmans RJ. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int J Climatol 2017;37:4302-15.

16. Harris I, Jones PD, Osborn TJ, Lister DH. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. Int J Climatol 2014;34:623-42.

17. Center for International Earth Science Information Network CCU. Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC); 2018.

18. van Doremalen N, Bushmaker T, Morris D, et al. Aerosol and surface stability of HCoV-19 (SARS-CoV-2) compared to SARS-CoV-1. medRxiv 2020;2020.03.09.20033217:doi: https://doi.org/10.1101/2020.03.09.20033217.

19. van Donkelaar A, Martin A, Brauer M, et al. Global Annual PM2.5 Grids from MODIS, MISR and SeaWiFS Aerosol Optical Depth (AOD) with GWR, 1998-2016. Palisades NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4ZK5DQS; 2018.

20. Hijmans RJ. raster: Geographic data analysis and modeling. R package version 3.0-7: http://CRAN.R-project.org/package=raster; 2019.

21. Legendre P, Legendre L. Numerical Ecology. Amsterdam: Elsevier; 2012.

22. Burnham KP, Anderson DR. Model selection and multimodel inference: a practical information-theoretic approach. New York: Springer Verlag; 2002.

23. Lukacs PM, Thompson WL, Kendall WL, et al. Concerns regarding a call for pluralism of information theory and hypothesis testing. J Appl Ecol 2007;44:456-60.

24. Giam XL, Olden JD. Quantifying variable importance in a multimodel inference framework. Methods Ecol Evol 2016;7:388-97.

25. Dormann CF, Elith J, Bacher S, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 2013;36:27-46.

26. Bates D, Maechler M, Bolker BM, Walker S. Fitting Linear Mixed-Effects Models Using lme4. J Stat Softw 2015;67:1-48.

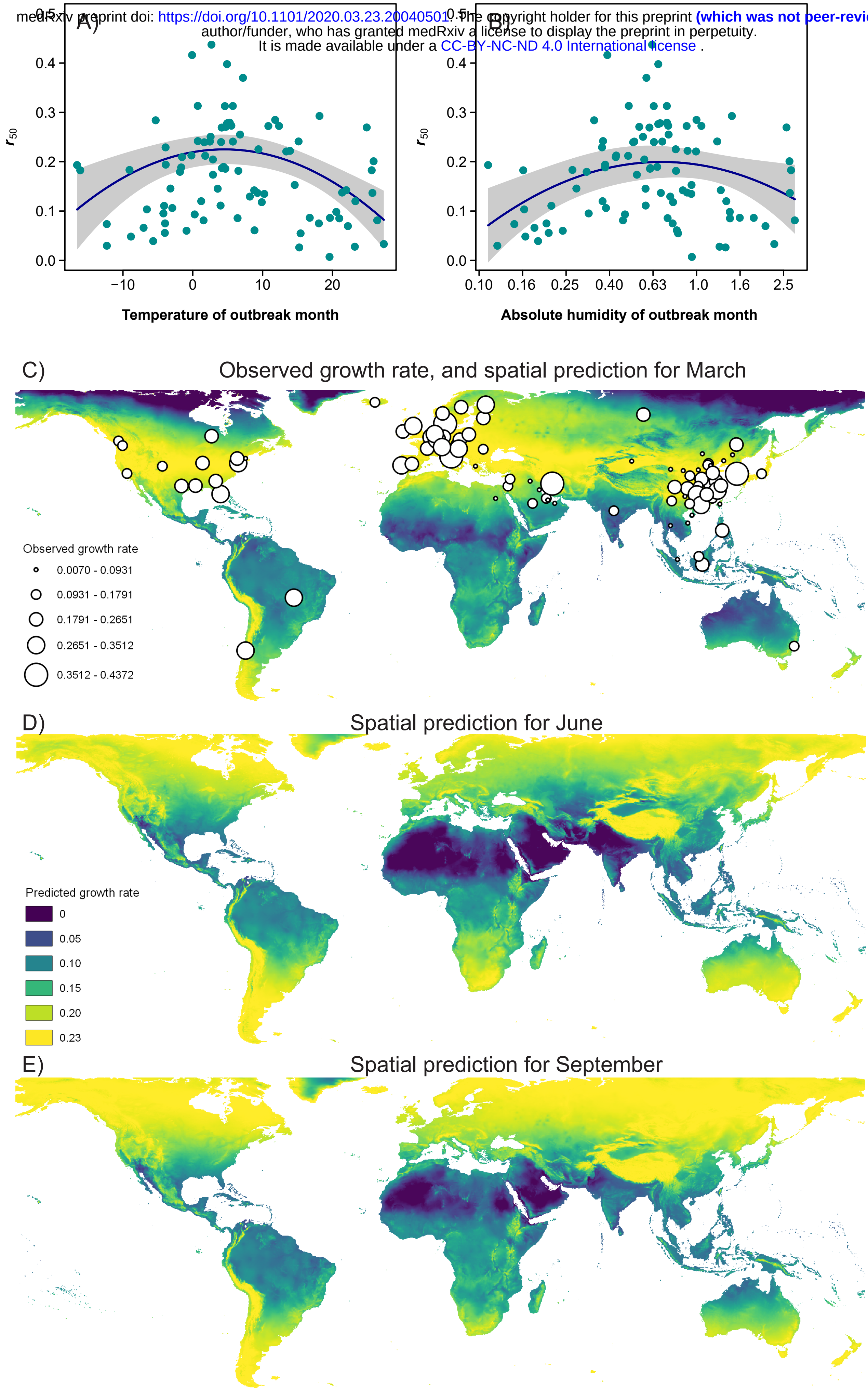
27. Kuznetsova A, Brockhoff B, Christensen HB. lmerTest Package: Tests in Linear Mixed Effects Models. J Stat Softw 2017;82:1-26.

28. Roser LG, Ferreyra LI, Saidman BO, Vilardi JC. EcoGenetics: An R package for the management and exploratory analysis of spatial data in landscape genetics. Mol Ecol Resour 2017;17:e241-e50.

29. Beale CM, Lennon JJ, Yearsley JM, Brewer MJ, Elston DA. Regression analysis of spatial data. Ecol Lett 2010;13:246-64.

Figure 1. Variation of Covid-19 growth rates in relation to climate, and spatial predictions for different months.

Variation of confirmed Covid-19 cases growth rates for the first 5 days after reaching a minimum threshold of 50 cases (r_{50}) during the January-March 2020 pandemic outbreak (n = 90 countries/regions, see list in Table S6) in relation to the mean temperature (**Panel A**) and to the mean absolute humidity of the outbreak month (**Panel B**). The lines are obtained from the best-fitting linear mixed models (LMMs) of r_{50} in relation to temperature or humidity, respectively (see Tables S3 and S5). The quadratic terms of both temperature and humidity were highly significant (temperature: $F_{1,87} = 14.4$, $P < 0.001$; humidity: $F_{1,84} = 7.82$, $P = 0.006$; full details in Tables S3 and S5). Shaded areas are 95% confidence band. **Panel C** shows the global patterns of r_{50}, with the size of dots is proportional to the observed r_{50} value. The background shows the spatial prediction of growth rates according to mean March temperatures[15]. Predictions are based on the best-fitting LMM of r_{50} in relation to mean temperature of the outbreak month (Table S3). **Panels D and E** show the spatial prediction of growth rates according to mean June and September temperatures[15], highlighting that optimal conditions for disease spread appear in temperate regions of the Southern Hemisphere.


A)
B)
r_{50}
0.0
0.1
0.2
0.3
0.4
0.5
−10
0
10
20
Temperature of outbreak month
0.10
0.16
0.25
0.40
0.63
1.0
1.6
2.5
Absolute humidity of outbreak month
C)
Observed growth rate, and spatial prediction for March
Observed growth rate
0.0070 - 0.0931
0.0931 - 0.1791
0.1791 - 0.2651
0.2651 - 0.3512
0.3512 - 0.4372
D)
Spatial prediction for June
Predicted growth rate
0
0.05
0.10
0.15
0.20
0.23
E)
Spatial prediction for September