Ecological Study on COVID-19: associations between the early exponential growth rate and historical environmental and socio-economic factors in 96 countries using GAM (Generalized Additive models)

Author: Thibault FIOLET, MSc. in Public Health and epidemiology E-mail: Thibault.fiolet@gmail.com

Please quote : *Fiolet Thibault. Ecological Study on COVID-19: associations between the early exponential growth rate and historical environmental and socio-economic factors in 96 countries using GAM (Generalized Additive models)*

Competing interests: The author declares that he has no competing interests. **Funding source**: No funding source has been received for this study. All datasets used in this study are openaccess.

Keywords: COVID-19; SRAS-CoV-2 ; air pollution ; temperature ; socio-economic determinants

Highlights:

- First ecological study to include socio-economical, environmental and demographic factors regarding to COVID-19 transmission in 96 countries
- The number of air passenger is the main driver of the early exponential growth rate
- Historical temperatures have a small significant inverse effect on transmission

Abstract:

SRAS-CoV-2 virus provoked a major pandemic. Some environmental parameters can influence the transmission of respiratory viruses, creating seasonal effects. Emerging viruses have a different behaviour since population immunity does not exist. Few studies looked at the determinants of the COVID-19 spread. This study investigates the link between historical socio-economic, environmental and health indicators and the transmission of COVID-19 in 96 countries using Generalized Additive Models. Our results indicated that the exponential early growth rate calculated on a period of seven days since the 100th case in each country is significantly and positively associated with the number of air passengers carried ($p<0.0001$). This spread was also associated with historical winter temperatures (p=0.02), the Universal Health Coverage Index (p=0.006) from World health Organization, GDP per capita (p=0.03) and the prevalence of overweight (=0.004) and undernourishment (p=0.09). In the sensitivity analysis, these results remained similar for a period of 14 days and with a growth rate since the $50th$ case. In conclusion, environmental effects on COVID-19 early transmission in this ecological study seem to be small.

Graphic Abstract

Main text:

On December 31, 2019, World Health Organization (WHO) identified in Wuhan (China) an unknown pneumonia. This new coronavirus rapidly spread around the world and on the $11th$ of March, the WHO declared pandemic. Most of countries closed schools, imposed social distancing measures and lockdown to flatten the epidemic curve (Viner et al., 2020). Beyond the different lag of COVID-19 onset in each country, some regions seemed to be less affected by this epidemic, which may not be only explained by health policies.

Respiratory viruses (pneumococcal infections, influenza) are often seasonal (Chew et al., 1998; Dowell et al., 2003; Visseaux et al., 2017). However, new emergent pathogens are less predictable. SRAS-CoV-1 appeared in 2003 and disappeared after only two years whereas MERS-CoV had a seasonal occurrence (Nassar et al., 2018). In laboratory conditions, SRAS-CoV and MERS-CoV are less stable at high temperature (Chan et al., 2011; Lm et al., 2010; van Doremalen et al., 2013). Understanding climate, socioeconomic and migrations determinants is crucial to managing an epidemic.

Some recent studies identified that population flow and air passenger volume modulate the SRAS-CoV-2 outbreaks (Jia et al., 2020; Lau et al., 2020). Comorbidities were present in half of the confirmed cases and were also associated with the disease severity (Richardson et al., 2020; Team, 2020; Zhou et al., 2020). Air pollution (PM2.5 and PM10 particles, nitrous oxide) was associated with COVID-19 incidence in China (Zhu et al., 2020) and fatality in Italy (Ogen, 2020). Temperature and humidity were also associated with the daily confirmed cases rate in China (Liu et al., 2020; Qi et al., 2020; Shi et al., 2020) and in Indonesia (Tosepu et al., 2020). Very few studies have analysed different risk factors in multiple countries.

The aim of this study was to identify potential environmental, socio-economic and health historical determinants of COVID-19 spread in the world, characterized by the growth rates of SRAS-CoV-2 through 96 countries. It is the first study to consider additional socio-economic factors which are potential confounding variables in the relation between environmental determinants and COVID-19 early growth rate.

Material and Methods

COVID-19 early exponential growth rate

Data on COVID-19 confirmed cases was retrieved from John Hopkins University Coronavirus Resource Center repository, as of April, 15 (CSSEGISandData, 2020).

In an early epidemic, the evolution of the cases can be approximated by an exponential model described by $N(t)=N_0 e^{rt}$ where N is the number of confirmed cases at time t

r represents the intrinsic exponential growth rate. N_0 is the number of cases at the initial time

r is preferred rather than the number of confirmed cases to assess the early spreading of epidemic because it does not depend on the onset time of the epidemic in each country.

The following adjusting factors were extracted for the most recent year available, from World Bank Open Data (WBOD) and World Health Organization (WHO) ("World Bank Open Data | Data," n.d.)

Environmental factors

Temperatures from 2-degree gridded data to the country were annually averaged for January-March period from 1961 to 1999.

Population-weighted PM2.5 air pollution was retrieved from WHO as a mean annual exposure during the 2015-2017 period.

Socio-economic and air traffic factors

These factors were retrieved from WBOD. Gross domestic product (GDP) per capita in 2017 and World Bank class income were used as an indicator for the economic development of countries. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. GDP per capita was missing for Venezuela. This missing value was imputed by a linear regression on the other health and populational indicators with *simputation* R package.

Air transport, as the number of passengers carried in 2018, include both domestic and international flights registered in the country.

Health factors

UHC service coverage index from WHO was used to assess the health care system performance. This index varied from 0 to 100 and is computed as the geometric mean of 14 indicators of essential health services (maternal, newborn and child health, infectious diseases, noncommunicable diseases and service capacity and access) ("GHO | By theme," n.d.).

The proportion of old people (over 65 years old) in percentage and population density were retrieved from WBOD. Population density is midyear population divided by land area in square kilometers. The prevalence of overweight adult based on the body mass index in 2016 and the prevalence of undernourishment (population below minimum level of dietary energy consumption) were retrieved from WHO. $ln(N_{\text{iour 7}} / N_{\text{iour 1}})/7$

Statistical analysis

Univariate analysis was performed using the Spearman rank test. Multivariate analyses were based on GAM modelling. The relationship between the growth rate and socio-economic and environmental determinants is modelled using a penalized GLM Generalized Additive Model (GAM). This model handles the complex nonlinearity of environmental variables (Ravindra et al., 2019). The general equation of this model is:

 $g(E[Y]) = \mu + f_1(X_1) + f_2(X_2) + ... + f_k(X_k)$

g() is a logarithm or identity function, f() are spline functions and μ is the intercept

Only 96 countries with more than 100 cases are included in this study. Growth rate is empirically calculated on the period of seven days since the 100th confirmed case.

$$
r = \frac{\ln\left(\frac{N(t = day\tau)}{N0(t = day\,1)}\right)}{7}
$$

Analysis of residuals for GAM models were plotted using the residuals function. Assumptions (gaussian assumption of residuals and constant variance) were checked with gam.check function in the *mgcv* R package.

Model 1 includes significant covariates from the univariable step with the Spearman when p<0.20. Model 2 introduces further adjustments on demographic factors.

Sensitivity analysis to test the robustness of the results was performed using a growth rate since the $50th$ case for a period of 7 days and using a growth rate since the 100th case for a period of 14 days. All analyses were performed using the *mgcv* package in R Studio statistical software (version 1.2.5001). The statistical tests were two-sided, and $p<0.05$ was considered statistically significant.

Results

Growth rate since the $100th$ case was significantly associated with annual winter temperatures (rho=-0.37), air passengers (rho=0.54), GDP per capita (rho=0.38), UHC index (rho=0.42), obesity (rho=0.42), under-nutrition (rho=-0.68). Iran (r=0.40), Turkey (r=0.36), Korea (r=0.35), Canada (r=0.34) and Ecuador (r=0.33) had the highest growth rate whereas Trinidad and Tobago (r=0.01), Cambodia $(r=0.02)$, Rwanda (r=0.02), Sri Lanka (r=0.03), Brunei (r=0.03) in figure 1. Mean and median growth rate are equal to r=0.15 (Table 1).

Table 1: Descriptive statistics for socio-economic, environmental, health and demographic factors

In multivariate GAM models, the early growth rate of COVID-19 was significantly positively associated with the number of air passengers ($p<0.0001$), winter temperatures ($p=0.02$), GDP per capita ($p=0.03$), UHC Index (p=0.006), prevalence of overweight (p=0.004) and under nutrition (p=0.09) in model 1 (Table 2). In figure 2, the economic determinants, the health coverage index, the prevalence of overweight and the temperature have a non-linear relationship. When the model was further adjusted with demographic factors in model 2, all of these previous determinants remain significant except for undernutrition ($p=0.10$). Population density and the proportion of population above 65 years old were not associated with the speed of COVID-19 early spread.

Table 2: parameter estimates for the GAM models for the growth rate on a period of seven days since 100th case

Model 1: adjusted on air passengers, temperature, air pollution, GDP per capita, UHC Index, prevalence of overweight and prevalence of undernourishment

Model 2: Model 1+ adjusted on population density and proportion of old people

Spline functions used in GAM model for each effects

In the sensitivity analysis, when the growth rate was calculated on a period of 14 days, air passengers, temperature, UHC Index and overweight effects remained significant. The results were similar for a growth rate calculated since the $50th$ case and a period of seven days.

Discussion

This study identified that the main drivers for the early COVID-19 epidemic growth rate was air traffic. Additional significant factors included the UHC index, prevalence of overweight and GDP per capita. The explained deviance (a pseudo-R2) in adjusted models increased to 52.9%. The number of air passengers was highly significant in each model and sensitivity analysis. More traffic is associated with a higher potential contacts rate, leading to a higher probability of transmission. Surprisingly, obesity effect was significant, it may be explained by a high correlation with high income countries which were more affected by the epidemic. The temperature effect is small.

Another ecological study on 44 countries found similar results with an OLS regression $(R²=34,5%)$ that the global transportation networks could be one of the main driver for the early transmission (Coelho et al., 2020). Four studies (Kassem, 2020; Liu et al., 2020; Notari, 2020; Shi et al., 2020) identified an inverse trend between temperature and the transmission of COVID-19 in China, in 24 and 42 countries. In the article of Notari, the linear regression had a very low \mathbb{R}^2 of 0,196 showing that the temperature effect is small.

In contrast, some studies found no link between meteorological factors and COVID-19 transmission in 224 Chinese cities (Yao et al., 2020). To finish, a short review identified the level of evidence to be low regarding to the temperature and humidity effects on the spread of COVID-19 (Mecenas et al., 2020). These inconsistencies may be explained by the heterogeneity in the locations or countries included, the scale (cities, regions or countries), in the outcome used (confirmed cases, standardized incidence ratio, number of death or growth rate).

The main pathways of SRAS-COV-2 transmission are direct contacts, indirect contacts, droplets and aerosol transmission is suspected (ref). Meteorological factors may have an impact on airborne transmission. Some animal studies tried to unravel the temperature and humidity factors. In guinea pigs, influenza transmission occurred more frequently at cold temperatures (Lowen et al., 2007). In lab conditions, SRAS-CoV-2 droplet on a surface was more stable at 4°C and sensitive to temperature (Chin et al., 2020). An Italian study identified a link between air pollution and a higher lethality for COVID-19 but in our study, we did not find an impact on the transmission (Conticini et al., 2020)

The strength of this analysis is the importance number of included countries. The significant determinants (mainly air traffic) are associated with a 50% of the variation on growth rate. Additionally, for the first time, historical socioeconomic, air traffic, environmental and demographic factors were added together in the models to study the early transmission of COVID-19.

This study has several limitations. In an ecological analysis, relationships cannot be extended at the individual levels. Second, the spread of COVID-19 is assessed by the growth rate but this parameter also relies on the number of confirmed cases. Testing capacity can influence the number of identified cases but it is not possible to have accurate statistics for testing capacities for every country. Residual confusion cannot be excluded. Lockdown and social distancing measures were not included because the growth rate was calculated in the onset of epidemic in each country. This study does not cover all countries in the world because it is impossible to have complete global dataset. Lastly, the environmental factors are based on historical data.

Conclusion

In conclusion, environmental factors (temperature, precipitation, PM2.5 particles) have a small effect regarding to the number of air passengers carried to explain the differences in the early growth rates at the global scale. Air traffic may be a main driver of this pandemic.

References

- Chan, K.H., Peiris, J.S.M., Lam, S.Y., Poon, L.L.M., Yuen, K.Y., Seto, W.H., 2011. The Effects of Temperature and Relative Humidity on the Viability of the SARS Coronavirus. Adv. Virol. 2011, 734690. https://doi.org/10.1155/2011/734690
- Chew, F.T., Doraisingham, S., Ling, A.E., Kumarasinghe, G., Lee, B.W., 1998. Seasonal trends of viral respiratory tract infections in the tropics. Epidemiol. Infect. 121, 121–128.
- Chin, A.W.H., Chu, J.T.S., Perera, M.R.A., Hui, K.P.Y., Yen, H.-L., Chan, M.C.W., Peiris, M., Poon, L.L.M., 2020. Stability of SARS-CoV-2 in different environmental conditions. Lancet Microbe 0. https://doi.org/10.1016/S2666-5247(20)30003-3
- Coelho, M.T.P., Rodrigues, J.F.M., Medina, A.M., Scalco, P., Terribile, L.C., Vilela, B., Diniz-Filho, J.A.F., Dobrovolski, R., 2020. Exponential phase of covid19 expansion is not driven by climate at global scale (preprint). Epidemiology. https://doi.org/10.1101/2020.04.02.20050773
- Conticini, E., Frediani, B., Caro, D., 2020. Can atmospheric pollution be considered a co-factor in extremely high level of SARS-CoV-2 lethality in Northern Italy? Environ. Pollut. 114465. https://doi.org/10.1016/j.envpol.2020.114465
- CSSEGISandData, 2020. CSSEGISandData/COVID-19.
- Dowell, S.F., Whitney, C.G., Wright, C., Rose, C.E., Schuchat, A., 2003. Seasonal Patterns of Invasive Pneumococcal Disease. Emerg. Infect. Dis. 9, 574–579. https://doi.org/10.3201/eid0905.020556
- GHO | By theme [WWW Document], n.d. . WHO. URL https://apps.who.int/gho/data/node.home (accessed 5.2.20).
- Jia, J.S., Lu, X., Yuan, Y., Xu, G., Jia, J., Christakis, N.A., 2020. Population flow drives spatiotemporal distribution of COVID-19 in China. Nature 1–11. https://doi.org/10.1038/s41586- 020-2284-y
- Kassem, A.Z.E., 2020. Do Weather Temperature and Median-age affect COVID-19 Transmission? (preprint). Infectious Diseases (except HIV/AIDS). https://doi.org/10.1101/2020.04.16.20067355
- Lau, H., Khosrawipour, V., Kocbach, P., Mikolajczyk, A., Ichii, H., Zacharski, M., Bania, J., Khosrawipour, T., 2020. The association between international and domestic air traffic and the coronavirus (COVID-19) outbreak. J. Microbiol. Immunol. Infect. https://doi.org/10.1016/j.jmii.2020.03.026
- Liu, J., Zhou, J., Yao, J., Zhang, X., Li, L., Xu, X., He, X., Wang, B., Fu, S., Niu, T., Yan, J., Shi, Y., Ren, X., Niu, J., Zhu, W., Li, S., Luo, B., Zhang, K., 2020. Impact of meteorological factors on the COVID-19 transmission: A multi-city study in China. Sci. Total Environ. 726, 138513. https://doi.org/10.1016/j.scitotenv.2020.138513
- Lm, C., S, J., Wa, R., Dj, W., Md, S., 2010. Effects of air temperature and relative humidity on coronavirus survival on surfaces. Appl. Environ. Microbiol. 76, 2712–2717. https://doi.org/10.1128/AEM.02291-09
- Lowen, A.C., Mubareka, S., Steel, J., Palese, P., 2007. Influenza Virus Transmission Is Dependent on Relative Humidity and Temperature. PLOS Pathog. 3, e151. https://doi.org/10.1371/journal.ppat.0030151
- Mecenas, P., Bastos, R., Vallinoto, A., Normando, D., 2020. Effects of temperature and humidity on the spread of COVID-19: A systematic review. (preprint). Public and Global Health. https://doi.org/10.1101/2020.04.14.20064923
- Nassar, M.S., Bakhrebah, M.A., Meo, S.A., Alsuabeyl, M.S., Zaher, W.A., 2018. Global seasonal occurrence of middle east respiratory syndrome coronavirus (MERS-CoV) infection. Eur. Rev. Med. Pharmacol. Sci. 22, 3913–3918. https://doi.org/10.26355/eurrev_201806_15276
- Notari, A., 2020. Temperature dependence of COVID-19 transmission (preprint). Epidemiology. https://doi.org/10.1101/2020.03.26.20044529
- Ogen, Y., 2020. Assessing nitrogen dioxide (NO2) levels as a contributing factor to coronavirus (COVID-19) fatality. Sci. Total Environ. 726, 138605. https://doi.org/10.1016/j.scitotenv.2020.138605
- Qi, H., Xiao, S., Shi, R., Ward, M.P., Chen, Y., Tu, W., Su, Q., Wang, W., Wang, X., Zhang, Z., 2020. COVID-19 transmission in Mainland China is associated with temperature and humidity: A time-series analysis. Sci. Total Environ. 728, 138778. https://doi.org/10.1016/j.scitotenv.2020.138778
- Ravindra, K., Rattan, P., Mor, S., Aggarwal, A.N., 2019. Generalized additive models: Building evidence of air pollution, climate change and human health. Environ. Int. 132, 104987. https://doi.org/10.1016/j.envint.2019.104987
- Richardson, S., Hirsch, J.S., Narasimhan, M., Crawford, J.M., McGinn, T., Davidson, K.W., Barnaby, D.P., Becker, L.B., Chelico, J.D., Cohen, S.L., Cookingham, J., Coppa, K., Diefenbach, M.A., Dominello, A.J., Duer-Hefele, J., Falzon, L., Gitlin, J., Hajizadeh, N., Harvin, T.G., Hirschwerk, D.A., Kim, E.J., Kozel, Z.M., Marrast, L.M., Mogavero, J.N., Osorio, G.A., Qiu, M., Zanos, T.P., 2020. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA. https://doi.org/10.1001/jama.2020.6775
- Shi, P., Dong, Y., Yan, H., Zhao, C., Li, X., Liu, W., He, M., Tang, S., Xi, S., 2020. Impact of temperature on the dynamics of the COVID-19 outbreak in China. Sci. Total Environ. 728, 138890. https://doi.org/10.1016/j.scitotenv.2020.138890
- Team, T.N.C.P.E.R.E., 2020. The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19) — China, 2020. China CDC Wkly. 2, 113–122. https://doi.org/10.46234/ccdcw2020.032
- Tosepu, R., Gunawan, J., Effendy, D.S., Ahmad, L.O.A.I., Lestari, H., Bahar, H., Asfian, P., 2020. Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia. Sci. Total Environ. 725, 138436. https://doi.org/10.1016/j.scitotenv.2020.138436
- van Doremalen, N., Bushmaker, T., Munster, V.J., 2013. Stability of Middle East respiratory syndrome coronavirus (MERS-CoV) under different environmental conditions. Euro Surveill. Bull. Eur. Sur Mal. Transm. Eur. Commun. Dis. Bull. 18. https://doi.org/10.2807/1560- 7917.es2013.18.38.20590
- Viner, R.M., Russell, S.J., Croker, H., Packer, J., Ward, J., Stansfield, C., Mytton, O., Bonell, C., Booy, R., 2020. School closure and management practices during coronavirus outbreaks including COVID-19: a rapid systematic review. Lancet Child Adolesc. Health 4, 397–404. https://doi.org/10.1016/S2352-4642(20)30095-X
- Visseaux, B., Burdet, C., Voiriot, G., Lescure, F.-X., Chougar, T., Brugière, O., Crestani, B., Casalino, E., Charpentier, C., Descamps, D., Timsit, J.-F., Yazdanpanah, Y., Houhou-Fidouh, N., 2017. Prevalence of respiratory viruses among adults, by season, age, respiratory tract region and type of medical unit in Paris, France, from 2011 to 2016. PLoS ONE 12. https://doi.org/10.1371/journal.pone.0180888
- World Bank Open Data | Data [WWW Document], n.d. URL https://data.worldbank.org/ (accessed 5.2.20).
- Yao, Y., Pan, J., Liu, Z., Meng, X., Wang, Weidong, Kan, H., Wang, Weibing, 2020. No Association of COVID-19 transmission with temperature or UV radiation in Chinese cities. Eur. Respir. J. https://doi.org/10.1183/13993003.00517-2020
- Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., Guan, L., Wei, Y., Li, H., Wu, X., Xu, J., Tu, S., Zhang, Y., Chen, H., Cao, B., 2020. Clinical course

and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The Lancet 395, 1054–1062. https://doi.org/10.1016/S0140- 6736(20)30566-3

Zhu, Y., Xie, J., Huang, F., Cao, L., 2020. Association between short-term exposure to air pollution and COVID-19 infection: Evidence from China. Sci. Total Environ. 727, 138704. https://doi.org/10.1016/j.scitotenv.2020.138704