Relative humidity predicts day-to-day variations in COVID-19 cases in the city of Buenos Aires

Andrea L. Pineda Rojas[1], Sandra M. Cordo, Ramiro I. Saurral[1,3], Jose L. Jimenez[4], Linsey C. Marr[5] and Emilio Kropff*[6]

[1] CIMA, UMI-IFAECI/CNRS, FCEyN, Universidad de Buenos Aires – UBA/CONICET, Buenos Aires, Argentina.

2 Laboratorio de Virología, QB, FCEyN, Universidad de Buenos Aires – IQUIBICEN/CONICET, Buenos Aires, Argentina.

3 DCAO, FCEyN, Universidad de Buenos Aires, Buenos Aires, Argentina.

4 Dept. of Chemistry and CIRES, University of Colorado, Boulder, CO, USA

5 Civil and Environmental Engineering, Virginia Tech, Blacksburg, Virginia, USA.

6 Leloir Institute – IIBBA/CONICET, Buenos Aires, Argentina.

KEYWORDS COVID-19, Relative humidity, airborne transmission, meteorology

ABSTRACT: Possible links between the transmission of COVID-19 and meteorology have been investigated by comparing positive cases across geographical regions. Little is known, however, about the degree to which meteorological conditions drive the daily dynamics of COVID-19 spread at a given location. The main limitation is that individual waves of the disease are typically abrupt and eventful, making correlations somewhat anecdotal. In contrast, we here present a long-term case study for the city of Buenos Aires, which has suffered a single prolonged wave of spread during 2020, with most significant changes in policy and population behavior taking place before the main local outbreak. We found that humidity plays a prominent role in modulating the variation of COVID-19 positive cases through a negative-slope linear relationship, with an optimal lag of 9 days between the meteorological observation and the positive case report. This relationship is specific to winter months, when relative humidity predicts up to half of the variance in positive cases. Our results provide a tool to anticipate local surges in COVID-19 cases after events of low humidity. More generally, they add to accumulating evidence pointing to dry air as an important driver of global COVID-19 transmission.

INTRODUCTION

Increasing evidence points to aerosols as a main mode of transmission of COVID-19, mostly occurring in indoor environments[1, 2]. Factors such as temperature and humidity can influence the transmission of respiratory viruses, as suggested by studies in laboratory-controlled conditions[3, 4]. The stability of influenza and other viruses in suspended aerosols, mimicking airborne transmission conditions, increases with low temperature and humidity[5, 6]. The exchange of water and heat with the environment also affects the physics of the droplets, determining their fate through deposition or dispersion[7, 8]. In addition, low levels of humidity also affect the immune response of the host, as demonstrated in laboratory animals[9]. Experiments of airborne transmission between animals confirm the overall facilitatory role played by low humidity in this complex chain of events[10].

Since the start of the pandemic, several studies have assessed the relationship between the number of daily COVID-19 cases and meteorological conditions in different regions of the world [11-16]. Indoor air conditions are not generally monitored, but widely available outdoor meteorology can be used as a proxy of indoor conditions, albeit mediated by levels of heating and ventilation[3].

Variables such as humidity, temperature, solar radiation, precipitation or wind speed have been found to co-vary with positive cases. These studies have typically taken two approaches. The first is to compare the spread of the disease across geographical regions. In this kind of study, variability in meteorological conditions is attained by including enough spatially distant locations, but differences other than meteorological (e.g., mitigation policies) between locations are a confounding factor. The second approach is to correlate the rising phase (or bursts) of the outbreak on a single location with a number of meteorological variables that might have modulated it. The challenge in this kind of study relates to the fact that many meteorological variables have marked yearly variations, and correlations with an abrupt increase in the number of COVID-19 cases might be spurious, simply reflecting two parallel but independent trends. In addition, the dynamic of local mitigation policies and the response of the population represent confounding elements for this approach. Due to these shortcomings, the question of whether or not meteorology can influence COVID-19 spread is still open, with sensible arguments on both sides[17, 18].

Here we present a long-term study of the influence of meteorological variables on the spread of COVID-19 in the city of Buenos Aires (CBA), Argentina. CBA has a humid subtropical climate with four distinct seasons, and is a useful case study for several reasons. First, over 5% of its 2.9 million inhabitants have suffered from COVID-19 in 2020, as

confirmed by PCR tests, during a single long wave of the spread starting at the beginning of May and reaching its peak by the end of August (Figure 1A). Second, although the area of the city is considerable (209 km^2), it is located on a flat terrain, so that surface meteorological conditions can be assumed to be horizontally homogeneous[19]. Third, the most disruptive policies were enforced when there were still fewer than 50 daily cases (lockdown: March 20th; facemasks mandatory in public spaces including shops, etc.: April 15th, 2020) in the hope of flattening the curve. Once it was clear that cases were rising, these measures were renewed roughly every 15 days, effectively extending the lockdown until November 9th, 2020. This period included the whole winter (June to September). Although during this period the population developed a certain level of relaxation toward some of the guidelines, this change in attitude had a slow progression, minimizing its possible influence on day-to-day variations in positive cases.

MATERIALS AND METHODS

Data gathering

Meteorological data corresponding to the station located at the domestic airport (AEP; World Meteorological Organization station number 87582) from January to November, 2020 was obtained from public databases provided by the National Oceanic and Atmospheric Administration (NOAA, US)[20] and OGIMET (www.ogimet.com). We included as meteorological variables for our study the surface daily mean values of relative humidity (RH), temperature (T), wind speed (WS), pressure (P), precipitation (PP) and sky cover (SC), as well as the surface daily minimum (T_{min}) and maximum (T_{max}) values of hourly temperature.

Data on COVID-19 cases (SARS-CoV2 positive PCR tests) in CBA for the same period were obtained from the daily reports by Health Ministry of Argentina, compiled and curated by Sistemas Mapache (github.com/SistemasMapache/Covid19arData). Data on the date in which symptoms began, available for only 60% of positive cases, were obtained from the Health Ministry of Argentina open data webpage (datos.salud.gob.ar/dataset). All data were processed through custom scripts written in MATLAB (MathWorks, Natick, MA).

Pre-processing

In order to study the mid-range dynamics of variability of new COVID-19 positive cases in CBA, avoiding the slow (wave of the pandemic) and fast (day-of-week effects) modulations, we defined the weekly difference of any given variable *X(t)* as the value taken by this variable on day *t* minus its value 7 days before:

$$\Delta X(t) = X(t) - X(t - 7 \text{ days}). \qquad (1)$$

Equation (1) was applied to daily COVID-19 positive cases to obtain *ΔCovid(t)* and to its meteorological counterparts (e.g. *ΔRH(t)* for relative humidity). To study the relationship between variables we used linear models and the Pearson correlation coefficient, including a potential lag between the day of the meteorological observation and that of the COVID-19 positive case report. Significance for a linear relationship against the null hypothesis of no contribution of a variable was obtained as a p-value through a t-test using the MATLAB function *fitlim()*. The optimal lag for a given variable was defined as the lag that maximized the significance and the absolute value of its correlation with *ΔCovid(t)*.

Note that the *Δ* transformation introduced in Equation 1, while convenient for a number of reasons, has a small disadvantage. Since a given value of *RH(t)* impacts on both *ΔRH(t)* and *ΔRH(t+7)*, a strong correlation at a given lag is typically mirrored by moderately strong correlations of opposite signs at lags 7 days earlier or later. To avoid any ambiguity, in all cases the optimal lag was determined by the global maximum of significance, regardless of other local maxima. Furthermore, while the *Δ* transformation is central to our results, an independent analysis that did not use it (see section 'event-triggered averages' below), was applied to the same data to confirm the typical size of optimal lags.

Optimization of the multivariate linear model

Full 10-fold cross-validation of all possible combinations of the 8 meteorological variables (all linear models with a number of variables ranging from 0 to 8) was implemented. For each combination of variables, the data was divided into 10 random subgroups of equal size. Each subgroup was once set aside and used as a test set for the linear model trained with the remaining 9 groups. The one-standard-error criterion was used to choose the best model, defined as the one with fewest variables that had a test sum of squared residuals not higher than the lowest one plus one standard error [21]. All linear combinations of variables were tested, including up to first order interactions, for two variations of the data: a) each variable with its own optimal lag and b) all variables using the optimal lag for *ΔRH*.

To corroborate the results we repeated the analyses using the MATLAB *lasso()* function, which implements the choice of the optimal model by penalizing the sum of the absolute value of linear coefficients under the one-standard-error criterion[21].

Event-triggered averages

To study the evolution of meteorological variables relative to peaks of *ΔCovid(t)* or the evolution of COVID-19 positive cases relative to extreme events of RH, the relevant events were identified, temporal windows of regular size defined around them, and variables averaged by overlapping these temporal windows.

The 7-day average of COVID-19 positive cases was used to understand the evolution of cases relative to extreme RH events along these short temporal windows. This evolution was computed as a percentage relative to the value on the day of the event by normalizing the 7-day average for each window by its value on that day.

RESULTS AND DISCUSSION

The curve of daily new COVID-19 cases in CBA exhibits a rich dynamic with different scales of temporal evolution (Figure 1A). As in many such datasets, fast and slow dynamics are observed. The slow dynamic reflects waves of spread of the pandemic, which in the particular case of CBA took during 2020 the form of a single, slowly modulated wave starting in May and reaching its peak in August. Day-of-week fluctuations (especially weekend vs. weekday differences) drive the fast dynamic, reflecting a multiplicity of factors that presumably range from the number of potentially dangerous social encounters to variations in testing and bureaucratic processing of the data. In addition, a third, intermediate dynamic range is

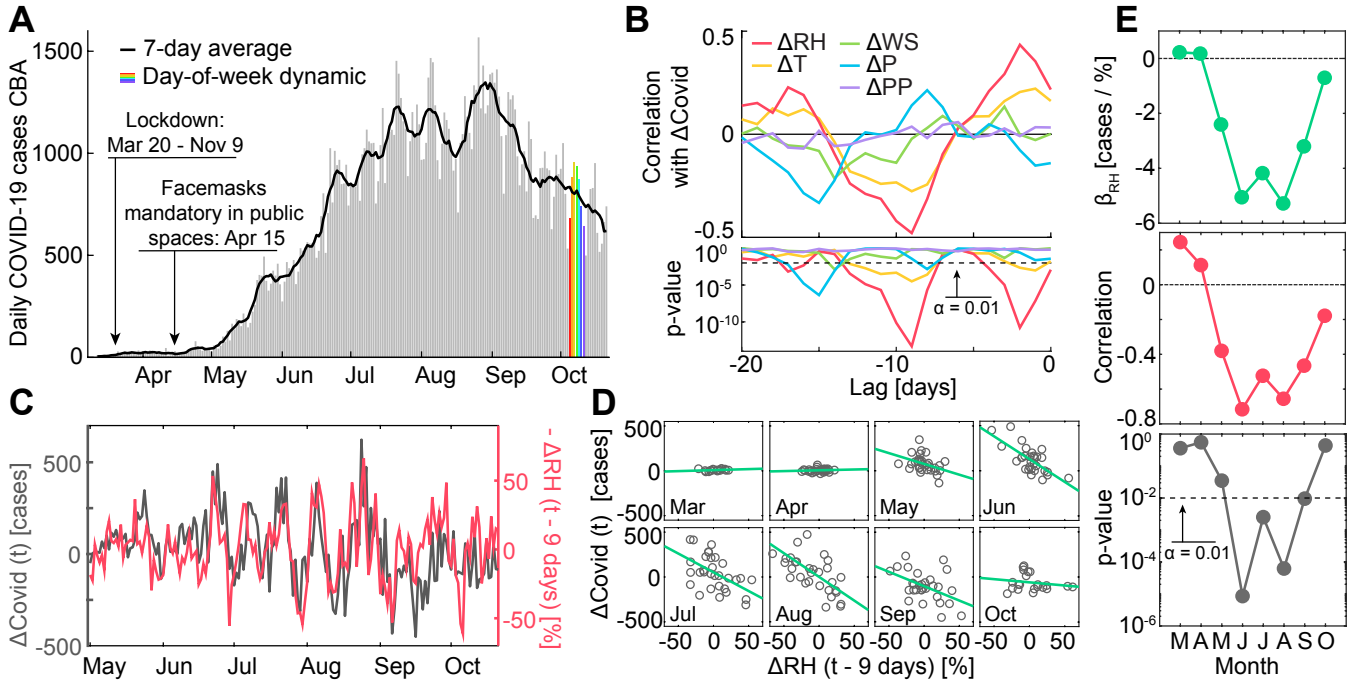

Figure 1. Relative humidity explains up to 50% of the variance in daily positive COVID-19 cases. (A) Daily cases in CBA (grey bars), 7-day average (black) and an example of the fast day-of-week dynamic (colors). Enforcement of major policy changes are indicated. (B) Top: Pearson correlation of *ΔCovid(t)* with *ΔRH(t+lag)*, for different lags in days, as well as with other meteorological variables (color coded). Bottom: Corresponding p-values (log scale). (C) Co-evolution of *ΔCovid(t)* (grey; left axis) and *ΔRH(t – 9 days)* (red; right axis). (D) Linear fits (green) for *ΔCovid(t)* vs *ΔRH(t – 9 days)* (grey) across individual months. (E) Slope (top), Pearson correlation (center) and corresponding p-values (bottom) for plots in (D).

apparent in this dataset, represented by fluctuations with an irregular periodicity in the range of 2-4 weeks. In the present work, we characterized these mid-range fluctuations following the hypothesis that they are at least in part rooted on meteorological conditions. To do so, we applied Equation 1 to all variables. Compared to other methods for filtering out day-of-week fluctuations, such as the 7-day average, this procedure has the advantage of allowing for instantaneous day-to-day fluctuations, without smoothing the variables along the temporal dimension.

We first studied the Pearson correlation between *ΔCovid(t)* and its meteorological counterparts (Figure 1B). Since meteorology is not expected to affect the outcome of positive COVID-19 cases reported on the same day, but rather those reported some days later, lags ranging from 0 to 20 days were included in the analysis. We found that *ΔRH*, *ΔT* and ΔT_{min} had a significant correlation ($p < 0.01$) with *ΔCovid* with an optimal lag of 9 days. The same happened for ΔT_{max} with a lag of 1 day, for *ΔWS* with a lag of 14 days and for *ΔP* with a lag of 15 days. As discussed in the Methods section, some variables exhibited a 7-day periodicity in the p-value curve, a direct consequence of Equation 1, but only the global minimum was considered to determine the optimal lag. Other variables did not exhibit a significant correlation with *ΔCovid(t)* regardless of the lag.

Of all variables, *ΔRH(t – 9days)* exhibited the most extreme correlation with *ΔCovid(t)* (Pearson correlation: -0.47; t(218): -7.9; p: 2×10^{-13}). The evolution of *ΔCovid(t)* and negative *ΔRH(t – 9 days)* was strikingly similar (Figure 1C), especially so when considering the numerous confounders expected to mediate a relationship between meteorology and the reporting of COVID-19 positive cases (individual variability in symptomatic response, test processing time or indoor vs. outdoor RH relationship to cite only a few).

To understand the relative importance of *ΔRH* compared to the other potential predictors, we applied two cross-validation methods (full 10-fold cross-validation and Lasso) to identify the linear combination of 8 meteorological variables (including individualized lags and first order interactions) that best modelled *ΔCovid(t)*. Both methods led to an identical conclusion, pointing to

$$\Delta\text{Covid}(t) = \beta_{constant} + \beta_{RH}\,\Delta\text{RH}(t - 9\text{ days}) \qquad (2)$$

as the best linear model of *ΔCovid(t)*, where β_X stands for the linear coefficient for variable *X*.

The optimal lag of 9 days includes time for the development of symptoms, testing and bureaucratic processing of data. To dissect this time window, we analyzed a second database that included, for around 60% of confirmed positive cases, the date in which symptoms began. Defining *ΔCovidS* in terms of number of patients with symptoms triggered at a given date, we found that its optimal lag relative to *ΔRH* was of 5 days (Pearson correlation: -0.44; t(218): -7.3; p: 5×10^{-12}). Since 5 days is the average window for the development of COVID-19 symptoms[22], this points to a specific role played by humidity in transmission and/or early stages of the disease.

Given that in CBA a single wave of the pandemic extended across seasons, we next assessed whether or not a seasonal effect modulated the *ΔCovid* - *ΔRH* relationship. We

classified the data by month of the COVID-19 report date (t) from March to October, 2020. For each month, we plotted $\Delta Covid(t)$ vs. $\Delta RH(t - 9\ days)$ and obtained the individual slopes (linear coefficient β_{RH} in Equation 2) and correlation coefficients (Figure 1D). We observed that the modulation of $\Delta Covid$ by ΔRH was only highly significant during the winter months (June to August), with similar values of β_{RH}. For transition months (May and September) the relationship weakened toward not significant, regardless of whether the overall number of COVID-19 positive cases was low (as in April) or high (as in October). Given that cross-validation procedures showed no significant interactions, this seasonal effect cannot be explained by the meteorological variables considered here. Mean monthly temperature, however, correlated with β_{RH} across months (Pearson correlation: 0.88; t(6): 4.6; p: 4 x 10^{-3}). We speculate that rather than causality, this correlation could reflect indirect mechanisms such as seasonal changes in behavior that do not rely on day-to-day meteorology (such as habits regarding ventilation, the use of heating or outdoor vs. indoor gathering).

Our results could provide a tool to anticipate local surges in COVID-19 due to low *RH*. To demonstrate this, we first studied the average evolution of *RH* in anticipation of events of extreme $\Delta Covid$ (Figure 2A). For the 7 peaks in $\Delta Covid$ that were higher than 200 positive cases, a trough in *RH* extending roughly from days 15 to 5 prior to the peak and reaching an average value of 55% at its minimum, was observed. Inversely, we asked what was the evolution of the 7-day average of COVID-19 cases after extreme low values of RH took place (Figure 2B). We observed an increase in COVID-19 cases starting 5 to 7 days after troughs of RH of different magnitude, peaking on average at day 11. For the 5 lowest humidity events (< 40%) that took place since the start of the pandemic, this implied an average increase of more than 20% in the 7-day average of COVID-19 positive cases.

This study provides a fresh perspective on the day-to-day dynamic of the COVID-19 pandemic. Further efforts should be directed to understand which of our results can be replicated in other locations of the world and which are specific to the CBA outbreak (and why). In addition, information regarding the type of transmission in each case (close proximity vs shared-room scale), if available, could shed light on the mechanisms by which RH modulates the number of COVID-19 cases. In close proximity, the physics of exhaled aerosols is likely to play a predominant role, while in shared-room transmission the viral decay rate could be more important. If, instead, RH modulated transmission mainly through its impact on the respiratory immune system, no difference would be expected between these groups. Further development in this area would improve our understanding of not only COVID-19 but possibly other respiratory diseases that pose a threat to human health and welfare.

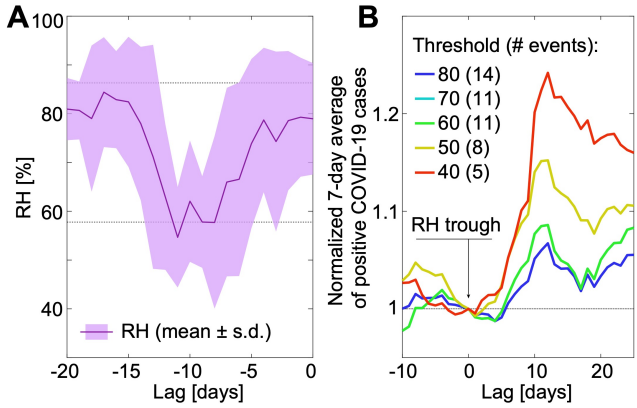

Figure 2. Extreme low events of RH anticipate surges in COVID-19 cases. (A) Distribution (mean ± s.d.) of RH during the 20 days prior to peaks in $\Delta Covid$ higher than 200 cases (7 events peaking at a lag = 0 days). Dashed lines: mean ± s.d. of RH for the whole dataset (March to October, 2020). (B) Mean evolution of the 7-day average of COVID-19 cases relative to extreme low values of RH (color coded threshold; number of events indicated). For every event, the 7-day average was normalized by its value at the trough of RH (lag = 0 days).

AUTHOR INFORMATION

Corresponding Author

* Emilio Kropff, Leloir Institute – IIBBA/CONICET, Av. Patricias Argentinas 435, CBA, Buenos Aires. Tel: +54 11 52387500. E-mail: ekropff@leloir.org.ar.

Author Contributions

The manuscript was written through contributions of all authors. / All authors have given approval to the final version of the manuscript.

Funding Sources

Notes

Any additional relevant notes should be placed here.

ACKNOWLEDGMENT

The work was supported by PICT 2015-1273 (E.K.), PICT 2018-1624 (A.P.R.) and PICT 2017-1136 (S.C.) grants from the Science Ministry of Argentina.

ABBREVIATIONS

CBA city of Buenos Aires; RH relative humidity; T temperature; WS wind speed; P pressure; PP precipitation.

REFERENCES

1. Prather, K. A.; Marr, L. C.; Schooley, R. T.; McDiarmid, M. A.; Wilson, M. E.; Milton, D. K., Airborne transmission of SARS-CoV-2. *Science* **2020,** *370* (6514), 303-304.
2. National Academies of Sciences, E.; Medicine, *Airborne Transmission of SARS-CoV-2: Proceedings of a*

Workshop—in Brief. The National Academies Press: Washington, DC, 2020; p 18.
3. Marr, L. C.; Tang, J. W.; Van Mullekom, J.; Lakdawala, S. S., Mechanistic insights into the effect of humidity on airborne influenza virus survival, transmission and incidence. *Journal of the Royal Society Interface* **2019,** *16* (150), 20180298.
4. Moriyama, M.; Hugentobler, W. J.; Iwasaki, A., Seasonality of respiratory viral infections. *Annual review of virology* **2020,** *7*.
5. Noti, J.; Blachere, F.; McMillen, C.; Lindsley, W.; Kashon, M., High Humidity Leads to Loss of Infectious Influenza Virus from Simulated. **2013**.
6. Yang, W.; Elankumaran, S.; Marr, L. C., Relationship between humidity and influenza A viability in droplets and implications for influenza's seasonality. *PloS one* **2012,** *7* (10), e46789.
7. Xie, X.; Li, Y.; Chwang, A.; Ho, P.; Seto, W., How far droplets can move in indoor environments-revisiting the Wells evaporation-falling curve. *Indoor air* **2007,** *17* (3), 211-225.
8. Bourouiba, L.; Dehandschoewercker, E.; Bush, J. W., Violent expiratory events: on coughing and sneezing. *Journal of Fluid Mechanics* **2014,** *745*, 537-563.
9. Kudo, E.; Song, E.; Yockey, L. J.; Rakib, T.; Wong, P. W.; Homer, R. J.; Iwasaki, A., Low ambient humidity impairs barrier function and innate resistance against influenza infection. *Proceedings of the National Academy of Sciences* **2019,** *116* (22), 10905-10910.
10. Lowen, A. C.; Mubareka, S.; Steel, J.; Palese, P., Influenza virus transmission is dependent on relative humidity and temperature. *PLoS Pathog* **2007,** *3* (10), e151.
11. Fu, S.; Wang, B.; Zhou, J.; Xu, X.; Liu, J.; Ma, Y.; Li, L.; He, X.; Li, S.; Niu, J., Meteorological Factors, Governmental Responses and COVID-19: Evidence from Four European countries. *Environmental research* **2020**, 110596.
12. Coccia, M., The effects of atmospheric stability with low wind speed and of air pollution on the accelerated transmission dynamics of COVID-19. *International Journal of Environmental Studies* **2020**, 1-27.
13. Gupta, A.; Banerjee, S.; Das, S., Significance of geographical factors to the COVID-19 outbreak in India. *Modeling earth systems and environment* **2020,** *6* (4), 2645-2653.
14. Rosario, D. K.; Mutz, Y. S.; Bernardes, P. C.; Conte-Junior, C. A., Relationship between COVID-19 and weather: Case study in a tropical country. *International journal of hygiene and environmental health* **2020,** *229*, 113587.
15. Ahmadi, M.; Sharifi, A.; Dorosti, S.; Ghoushchi, S. J.; Ghanbari, N., Investigation of effective climatology parameters on COVID-19 outbreak in Iran. *Science of the Total Environment* **2020**, 138705.
16. Briz-Redón, Á.; Serrano-Aroca, Á., The effect of climate on the spread of the COVID-19 pandemic: A review of findings, and statistical and modelling techniques. *Progress in Physical Geography: Earth and Environment* **2020,** *44* (5), 591-604.
17. Carlson, C. J.; Gomez, A. C.; Bansal, S.; Ryan, S. J., Misconceptions about weather and seasonality must not misguide COVID-19 response. *Nature Communications* **2020,** *11* (1), 1-4.
18. Coccia, M., Factors determining the diffusion of COVID-19 and suggested strategy to prevent future accelerated viral infectivity similar to COVID. *Science of the Total Environment* **2020**, 138474.
19. Pineda Rojas, A. L.; Borge, R.; Mazzeo, N. A.; Saurral, R. I.; Matarazzo, B. N.; Cordero, J. M.; Kropff, E., High PM10 concentrations in the city of Buenos Aires and their relationship with meteorological conditions. *Atmospheric Environment* **2020,** *241*, 117773.
20. Lott, N.; Vose, R.; Del Greco, S.; Ross, T.; Worley, S.; Comeaux, J. In *The integrated surface database: Partnerships and progress*, Extended Abstracts, 24th Conf. on Interactive Information and Processing Systems, 2008.
21. James, G.; Witten, D.; Hastie, T.; Tibshirani, R., *An introduction to statistical learning*. Springer: 2013; Vol. 112.
22. 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* **2020,** *25* (5), 2000062.