

Effect of Temperature on the Transmission of COVID-19: A Machine Learning Case Study in Spain

Amir Abdollahi^{a,*}, Maryam Rahbaralam^b

^a*Laboratori de Càlcul Numèric, Department of Civil and Environmental Engineering, Universitat Politècnica de Catalunya, 08034 Barcelona, Spain*

^b*Barcelona Supercomputing Center
08034 Barcelona, Spain*

Abstract

The novel coronavirus (COVID-19) has already spread to almost every country in the world and has infected over 3 million people. To understand the transmission mechanism of this highly contagious virus, it is necessary to study the potential factors, including meteorological conditions. Here, we present a machine learning approach to study the effect of temperature, humidity and wind speed on the number of infected people in the three most populous autonomous communities in Spain. We find that there is a moderate inverse correlation between temperature and the daily number of infections. This correlation manifests for temperatures recorded up to 6 days before the onset, which corresponds well to the known mean incubation period of COVID-19. We also show that the correlation for humidity and wind speed is not significant.

Keywords: Meteorological conditions, COVID-19, Novel coronavirus, Machine Learning, Transmission

1. Introduction

Coronavirus disease 2019 (COVID-19) is an infectious disease first reported in Wuhan, China, in late December 2019. This disease is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1, 2], which has spread rapidly around the world. In response to this outbreak, the World

*Corresponding author (amir.abdollahi@upc.edu)

Email address: amir.abdollahi@upc.edu (Amir Abdollahi)

Health Organization (WHO) characterised COVID-19 as a pandemic on 11 March [3]. As of today, the current COVID-19 outbreak situation has confirmed over 3 million cases and 208 thousand deaths globally [4]. Spain is one of the most affected countries with over 210 thousand confirmed cases and 23 thousand confirmed deaths [5] (data reported on April 28th 2020). COVID-19 infection could cause severe respiratory disorders with symptoms such as fever, coughing and shortness of breath. The average incubation period is 5–6 days with the longest incubation period of 14 days [6, 7, 8, 9, 10]. In addition to person-to-person transmission of COVID-19 through direct contact or droplets [11], studies have shown that geographical, and climatic factors are important factors in the transmission and survival of coronaviruses [12, 2, 13, 14, 15, 16, 17]. For example, temperature could increase or decrease the transmission risk by affecting the survival time of coronaviruses on surfaces [18, 19, 20, 21, 22]. Since the outbreak of COVID-19, many researchers have focused their efforts in understanding the influencing factors and transmission mechanism of this virus. These efforts are motivated by previous investigations on other viral infectious diseases such as SARS and Influenza. SARS is another coronavirus type that appeared in 2002 and caused a pandemic with more than 8 thousand infected people in 26 countries. Noticeable correlations between the SARS cases and the environmental temperature were reported [19, 23]. SARS transmission appeared to be dependent on seasonal temperature changes and the multiplicative effect of hospital infection [24]. It has also been documented that several factors, including climatic conditions (temperature and humidity), and population density might have affected the patterns of influenza spread and diversification [25].

Several studies have demonstrated that there is a correlation between meteorological factors, such as temperature and humidity, and the transmission of COVID-19 [26, 23]. Recent work has shown that there is a significant positive effect of the diurnal temperature range (DTR) on the daily mortality of COVID-19, and a significant negative association between COVID-19 mortality and ambient temperature as well as absolute humidity [27]. Tosepu and coworkers showed that only average temperature was significantly correlated with COVID-19 [28]. However, other studies have shown that temperature and humidity are not correlated with the transmission of COVID-19 [29, 30, 31]. This contradiction suggests a need for further investigation on this topic.

In this study, we explore the correlation between meteorological condi-

tions and the daily increment of infected people due to COVID-19 in Spain. Given that Spain is the most climatically diverse country in Europe [32] and has one of the highest numbers of infected people, it becomes a potential case study country for this purpose. First, we introduce the methods used in this paper, including the study area, data collection, study period, and data analysis methods, then we present our analysis and results. Finally, we conclude the primary outcomes of our study.

2. Methods

2.1. Case study area

We consider the three most populous autonomous communities Anadalucía, Cataluña, and Madrid in Spain for our case study. Table 1 shows the population of these three communities in 2019 [33] and the number of confirmed COVID-19 cases until April 28th, 2020 [34]. In addition to their high population, these communities are located in different parts of Spain with varying conditions of climate. Figure 1 shows the average temperature of the three communities during March 2020. A temperature variation of up to about 16 °C is observed between these communities, highlighting the climate diversity of Spain.

Community	Population	Confirmed cases
Anadalucía	8,446,561	13,250
Cataluña	7,609,499	49,185
Madrid	6,685,471	63,989

Table 1: Population in 2019 [33] and the number of confirmed COVID-19 cases [34] in the three most populous autonomous communities in Spain until April 28th, 2020.

2.2. Data collection

To study the effect of meteorological conditions on the transmission of COVID-19, here we consider the daily increment of infected people in the three communities. This dataset was extracted from Datadista Github repository [34], containing the updated daily information according to the calendar and publication of the Spanish Ministry of Health. For each daily record, we also obtain meteorological data such as the averages of temperature, humidity, and wind speed for the same day and the previous 13 days, leading to a total of 42 predictors. These data were extracted from Wunderground,

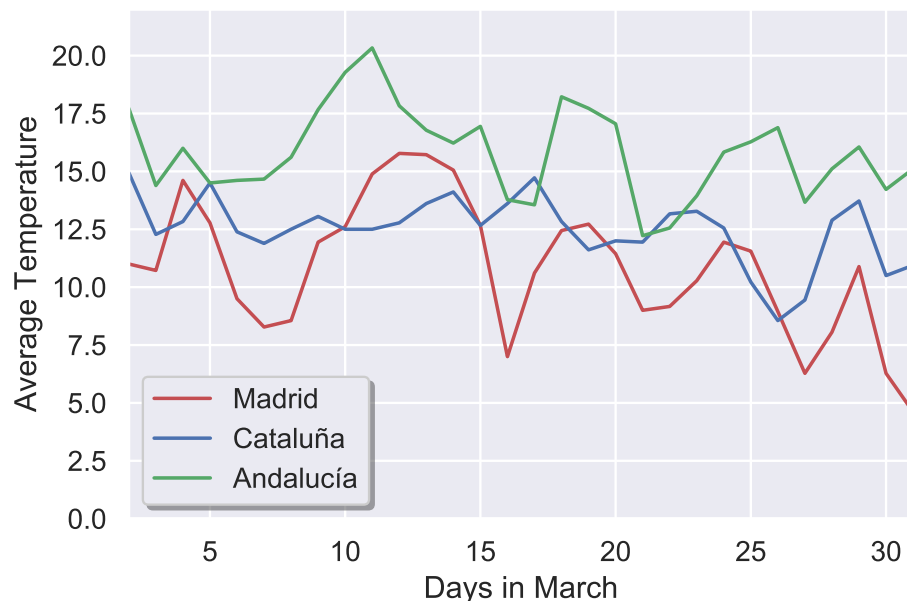


Figure 1: Average temperature during March 2020 in the three most populous autonomous communities in Spain; Madrid, Cataluña, and Andalucía.

an online open-source weather database [35]. After merging these predictors with the daily increment of infected people for each community, then we concatenate the three databases to form a single dataset for our analysis.

2.3. Case study period

Figure 2 shows the daily increment of infected people for the communities of Madrid, Cataluña, and Andalucía from March 1st until April 28th, 2020. All the three communities reached the peak of daily infections until the end of March with an upward trend during the month. A state of alarm and a national lockdown was imposed on March 14th [36], whose effect manifests as a downward trend in the number of daily cases from the beginning of April (day 31). To minimise the impacts of the lockdown on our analysis, we consider the upward transmission period during March, see Fig 2. Also, we focus our case study on the last 17 days of March to capture all the possible transmissions during the first 14 days, known as the maximum incubation period for the virus [8, 9, 10].

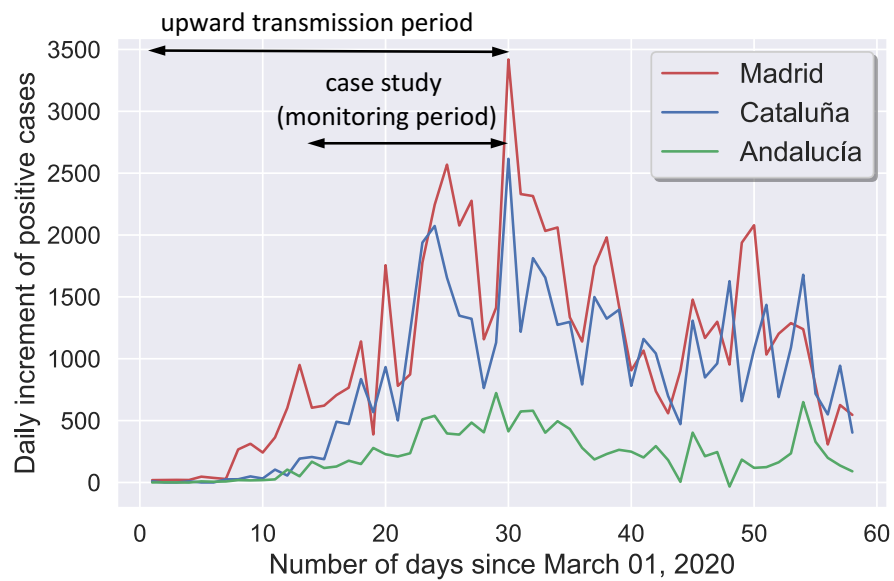


Figure 2: Daily increment of infected people as a function of the number of days since March 1st, 2020, for the three most populous autonomous communities in Spain; Madrid, Cataluña, and Andalucía. For our case study, we consider a monitoring period of 17 days when the number of infections shows an upward trend.

2.4. Data analysis

In this study, we use the Pearson's correlation coefficient as a measure of the linear association between the predictors and the target. Like other correlation coefficients, the Pearson's correlation varies between -1 and +1 with 0 implying no correlation. The p-value roughly indicates the probability of an uncorrelated system producing datasets that have a Pearson correlation at least as extreme as the one computed from these datasets. We also employ linear regression, a supervised machine learning technique, to find the relationship between the most important predictors and the target [37]. Here, we use Ordinary Least Squares (OLS) to estimate the unknown regression coefficients. A range of plausible values for these coefficients can be obtained by calculating confidence intervals.

3. Results and Discussion

Figure 3 presents the correlation between the meteorological predictors and the daily increment of infected people (target). It is interesting to observe that the averages of temperature up to the previous 6 days (T-AVG, and T-AVG-1 to -6) show a moderate negative correlation range between -0.7 to -0.5 (p-value $< 10^{-3}$). This result corresponds well with the mean incubation period of COVID-19, which was estimated to be 5-6 days [8, 9, 10]. We note that none of the 14 temperature predictors shows a positive correlation with the target. Humidity and wind speed do not indicate this consistency, and they also show a weak correlation in the range of -0.5 to 0.5.

We perform linear regression between the target and the 5-day lag temperature (T-AVG-5), which shows the highest correlation. Figure 4 shows the data points, the regression line, as well as the confidence interval (95%). Considering a linear regression model $I = \beta_0 + \beta_1 T$, I and T being the daily increment of infected people and the average temperature, respectively, an intercept of $\beta_0 = 3556$ (95% CI, 2813 to 4298) and a slope of $\beta_1 = -194$ $^{\circ}\text{C}^{-1}$ (95% CI, -248 to -140) are obtained from the linear regression results. In other words, a 1°C increase in average temperature leads to a decrease of 194 daily positive cases (95% CI, 140 to 248). We also perform linear regression considering the other important temperature predictors. Table 2 summarizes the regression coefficients. The coefficient of β_1 shows a range from -178 $^{\circ}\text{C}^{-1}$ to -136 $^{\circ}\text{C}^{-1}$, while the intercept β_0 varies from 2691 to 3556 among different predictors.

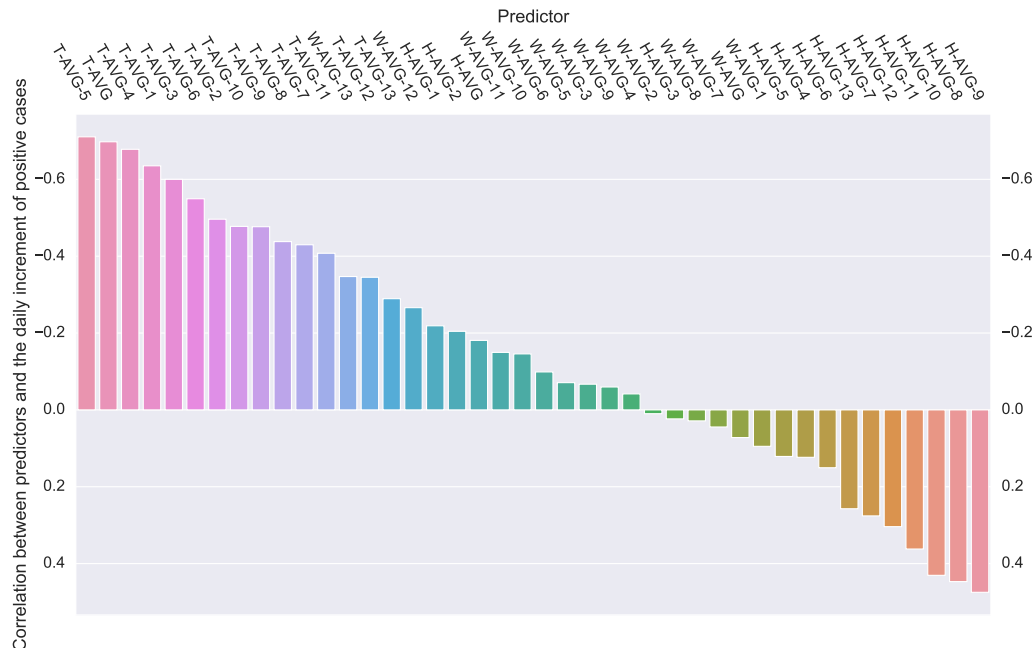


Figure 3: Correlation between the meteorological predictors and the daily increment of infected people. The averages of temperature, humidity and wind speed are represented by T-AVG, H-AVG, and W-AVG, respectively. The following number shows the number of days' lag between the date when these meteorological predictors were measured and the date when the number of infections is considered.

Temperature predictor	β_1	β_1 (CI)	β_0	β_0 (CI)
T-AVG	-178	(-229, -127)	3142	(2488,3796)
T-AVG-1	-170	(-228, -112)	3091	(2337,3845)
T-AVG-2	-136	(-202,-69)	2691	(1808,3574)
T-AVG-3	-156	(-215,-98)	2984	(2198,3770)
T-AVG-4	-176	(-229,-123)	3264	(2539,3988)
T-AVG-5	-194	(-248,-140)	3556	(2813,4298)
T-AVG-6	-155	(-221,-89)	3032	(2120,3945)

Table 2: Regression coefficients and 95% confidence interval (CI) for important temperature predictors.

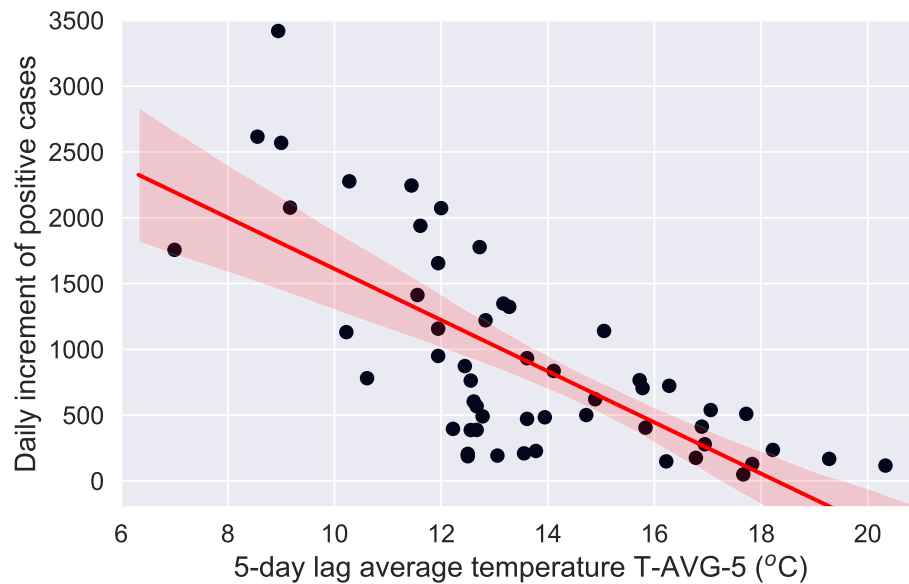


Figure 4: Daily increment of infected people as a function of the 5-day lag average temperature (T-AVG-5). The regression line and the confidence interval (95%) are highlighted in red.

4. Conclusions

We have presented a machine learning case study to evaluate the effect of meteorological factors, such as temperature, humidity, and wind speed on the transmission of the novel coronavirus (COVID-19). We have found that the average temperatures of up to 6 previous days have a moderate negative correlation with the daily increment of infected people. This result is relevant since the mean incubation period of COVID-19 is known to be around 5-6 days [8]. We have also performed a linear regression which has shown a decrease of about 200 daily infected people per 1°C increase in average temperature. We note that due to several uncertainties involved in testing and collecting the number of infected people, and other factors affecting the transmission, our analysis and results need to be interpreted with caution. However, our results could provide a clue to understanding the transmission mechanism of COVID-19.

Declaration of competing interest

The authors declare no conflict of interest.

References

- [1] A. E. Gorbalenya, S. C. Baker, R. S. Baric, R. J. de Groot, C. Drosten, A. A. Gulyaeva, B. L. Haagmans, C. Lauber, A. M. Leontovich, B. W. Neuman, D. Penzar, S. Perlman, L. L. Poon, D. Samborskiy, I. A. Sidorov, I. Sola, and J. Ziebuhr, “Severe acute respiratory syndrome-related coronavirus: The species and its viruses – a statement of the coronavirus study group,” *bioRxiv*, 2020.
- [2] S. Z. F. Wu, B. Yu, Y. Chen, W. Wang, Z. Song, Y. Hu, Z. Tao, J. Tian, Y. Pei, M. Yuan, Y. Zhang, F. Dai, Y. Liu, Q. Wang, J. Zheng, L. Xu, E. Holmes, and Y. Zhang, “A new coronavirus associated with human respiratory disease in china,” *Nature*, vol. 579, p. 265–269, 2020.
- [3] D. Cucinotta and M. Vanelli, “Who declares covid-19 a pandemic,” *Acta Bio Medica Atenei Parmensis*, vol. 91, pp. 157–160, Mar. 2020.
- [4] “Coronavirus Disease (COVID-19) Pandemic, World Health Organization, 2020.” <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>.
- [5] “Situación de COVID-19 o Coronavirus en España.” <https://covid19.isciii.es/>.
- [6] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, and B. Cao, “Clinical features of patients infected with 2019 novel coronavirus in wuhan, china,” *The Lancet*, vol. 395, no. 10223, pp. 497 – 506, 2020.
- [7] H. Nishiura, N. M. Linton, and A. R. Akhmetzhanov, “Serial interval of novel coronavirus (covid-19) infections,” *International Journal of Infectious Diseases*, vol. 93, pp. 284 – 286, 2020.
- [8] S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, and J. Lessler, “The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application,” *Annals of Internal Medicine*, 03 2020.

- [9] N. Linton, T. Kobayashi, Y. Yang, K. Hayashi, A. Akhmetzhanov, S.-M. Jung, B. Yuan, R. Kinoshita, and H. Nishiura, “Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: A statistical analysis of publicly available case data,” *Journal of clinical medicine*, vol. 9, 02 2020.
- [10] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T. T. Lam, J. T. Wu, G. F. Gao, B. J. Cowling, B. Yang, G. M. Leung, and Z. Feng, “Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia,” *New England Journal of Medicine*, vol. 382, no. 13, pp. 1199–1207, 2020. PMID: 31995857.
- [11] C.-C. Lai, T.-P. Shih, W.-C. Ko, H.-J. Tang, and P.-R. Hsueh, “Severe acute respiratory syndrome coronavirus 2 (sars-cov-2) and coronavirus disease-2019 (covid-19): The epidemic and the challenges,” *International Journal of Antimicrobial Agents*, vol. 55, no. 3, p. 105924, 2020.
- [12] A. Casadevall, “Climate change brings the specter of new infectious diseases,” *The Journal of Clinical Investigation*, vol. 130, pp. 553–555, 2 2020.
- [13] H. Qi, S. Xiao, R. Shi, M. P. Ward, Y. Chen, W. Tu, Q. Su, W. Wang, X. Wang, and Z. Zhang, “Covid-19 transmission in mainland china is associated with temperature and humidity: A time-series analysis,” *Science of The Total Environment*, vol. 728, p. 138778, 2020.
- [14] M. Şahin, “Impact of weather on covid-19 pandemic in turkey,” *Science of The Total Environment*, vol. 728, p. 138810, 2020.
- [15] F. A. Cássaro and L. F. Pires, “Can we predict the occurrence of covid-19 cases? considerations using a simple model of growth,” *Science of The Total Environment*, vol. 728, p. 138834, 2020.
- [16] M. Saez, A. Tobias, D. Varga, and M. A. Barceló, “Effectiveness of the measures to flatten the epidemic curve of covid-19. the case of spain,” *Science of The Total Environment*, vol. 727, p. 138761, 2020.

- [17] M. Coccia, “Factors determining the diffusion of covid-19 and suggested strategy to prevent future accelerated viral infectivity similar to covid,” *Science of The Total Environment*, p. 138474, 2020.
- [18] P. Bi, J. Wang, and J. E. Hiller, “Weather: driving force behind the transmission of severe acute respiratory syndrome in china?,” *Internal Medicine Journal*, vol. 37, no. 8, pp. 550–554, 2007.
- [19] J. Tan, L. Mu, J. Huang, S. Yu, B. Chen, and J. Yin, “An initial investigation of the association between the sars outbreak and weather: with the view of the environmental temperature and its variation,” *Journal of Epidemiology & Community Health*, vol. 59, no. 3, pp. 186–192, 2005.
- [20] L. M. Casanova, S. Jeon, W. A. Rutala, D. J. Weber, and M. D. Sobsey, “Effects of air temperature and relative humidity on coronavirus survival on surfaces,” *Applied and Environmental Microbiology*, vol. 76, no. 9, pp. 2712–2717, 2010.
- [21] K.H.Chan, J. S. M. Peiris, S. Y. Lam, L. L. M. Poon, K.Y.Yuen, and W. Seto, “The effects of temperature and relative humidity on the viability of the sars coronavirus,” *Advances in Virology*, 2011.
- [22] N. van Doremalen, T. Bushmaker, and V. J. Munster, “Stability of middle east respiratory syndrome coronavirus (mers-cov) under different environmental conditions,” *Eurosurveillance*, vol. 18, no. 38, 2013.
- [23] J. Yuan, H. Yun, W. Lan, W. Wang, S. G. Sullivan, S. Jia, and A. H. Bittles, “A climatologic investigation of the sars-cov outbreak in beijing, china,” *American Journal of Infection Control*, vol. 34, no. 4, pp. 234 – 236, 2006.
- [24] K. LIN, D. Y.-T. FONG, B. ZHU, and J. KARLBERG3, “Environmental factors on the sars epidemic: air temperature, passage of time and multiplicative effect of hospital infection,” *Epidemiology and infection*, vol. 134, pp. 223–30, 2006.
- [25] B. D. Dalziel, S. Kissler, J. R. Gog, C. Viboud, O. N. Bjørnstad, C. J. E. Metcalf, and B. T. Grenfell, “Urbanization and humidity shape the intensity of influenza epidemics in u.s. cities,” *Science*, vol. 362, no. 6410, pp. 75–79, 2018.

- [26] Y. Ma, Y. Zhao, J. Liu, X. He, B. Wang, S. Fu, J. Yan, J. Niu, J. Zhou, and B. Luo, “Effects of temperature variation and humidity on the death of covid-19 in wuhan, china,” *Science of The Total Environment*, vol. 724, p. 138226, 2020.
- [27] B. Oliveiros, L. Caramelo, N. C. Ferreira, and F. Caramelo, “Role of temperature and humidity in the modulation of the doubling time of covid-19 cases,” *medRxiv*, 2020.
- [28] R. Tosepu, J. Gunawan, D. S. Effendy, L. O. A. I. Ahmad, H. Lestari, H. Bahar, and P. Asfian, “Correlation between weather and covid-19 pandemic in jakarta, indonesia,” *Science of The Total Environment*, vol. 725, p. 138436, 2020.
- [29] P. Shi, Y. Dong, H. Yan, X. Li, C. Zhao, W. Liu, M. He, S. Tang, and S. Xi, “The impact of temperature and absolute humidity on the coronavirus disease 2019 (covid-19) outbreak - evidence from china,” *medRxiv*, 2020.
- [30] T. Jamil, I. S. Alam, T. Gojobori, and C. Duarte, “No evidence for temperature-dependence of the covid-19 epidemic,” *medRxiv*, 2020.
- [31] J. Xie and Y. Zhu, “Association between ambient temperature and covid-19 infection in 122 cities from china,” *Science of The Total Environment*, vol. 724, p. 138201, 2020.
- [32] <http://www.aemet.es/documentos/es/conocermas/publicaciones/Atlas-climatologico/Atlas.pdf>.
- [33] <https://www.statista.com/statistics/445549/population-of-spain-by-autonomous-community>.
- [34] <https://github.com/datadista/datasets/tree/master/COVID>
- [35] <https://www.wunderground.com>.
- [36] <https://administracion.gob.es/pagHome/atencionCiudadana/Estado-de-alarma-crisis-sanitaria.html.XqvzAS9h10t>.
- [37] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.