Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
The spread of 2019-nCoV in China was primarily driven by population density. Comment on “Association between short-term exposure to air pollution and COVID-19 infection: Evidence from China” by Zhu et al.

Keywords:
Novel coronavirus
2019-nCoV
COVID-19
Population density
Demography
Air pollution

1. Introduction and background

The outbreak of the coronavirus pandemic (Cheng and Shan, 2020) has stimulated a multitude of studies on the topic over just a few months. Yet a few of them go beyond the clinical scope and try to deal with other epidemiological aspects. However, both the earlier literature on other viruses and some recent studies on the novel coronavirus have examined the likely relationship between socio-economic and environmental conditions and the diffusion of pandemics. In particular, those studies point to the role potentially played by weather conditions (Iqbal et al., 2020; Sobral et al., 2020), transportation (Adda, 2016; Jia et al., 2020), economic activity (Sarmadi et al., 2020), and air pollution (Coccia, 2020; Conticini et al., 2020). As far as the latter aspect is concerned, it is worth noting that pollution emissions are already known to be associated with respiratory viral infections (Becker and Soukup, 1999; Cienczwicki and Jaspers, 2007; Cui et al., 2003; Horne et al., 2018; Mehta et al., 2013; Xu et al., 2016; Ye et al., 2016). Recently, an article published in the journal Science of the Total Environment has supported the “Association between short-term exposure to air pollution and COVID-19 infection” (Zhu et al., 2020) based on an analysis of 120 Chinese cities using a generalized additive model. The authors find “significantly positive associations of PM2.5, PM10, CO, NO2, and O3 with COVID-19 confirmed cases” (p. 3).

In this note, we show that the results of the study mentioned above may be affected by the issue of spurious correlation due to the omission of a common factor, namely, population density. The issue lies in the fact that, except for weather conditions, the concurrent factors suggested so far - i.e., transportation volumes, economic activity, and air pollution - are anthropic in nature. Thus, they all depend on the extent of human activities: the larger the population is, the higher the transportation volumes, economic activity, air pollution, and virus infections are (Fig. 1). Accordingly, when it comes to measuring the actual effect of anthropogenic causes on the pandemic, normalizing by population size and controlling for population density (Coccia, 2020) is by no means an option. Under the above framework, we consider alternative modeling, whose data covers almost all Chinese provinces and their socio-economic variables, relativized by the population size. Further, we also consider environmental variables that may explain the infection rate. Our main aim is to raise questions about future directions of the research on ecological and socio-economic aspects of 2019-nCoV.

2. Materials and method

2.1. Outline of the problem

To show how the issue of spurious correlation may affect the relationship between virus contamination (Y) and air pollution (X) though population (Pop), we consider data of people infected by COVID-19 and the level of sulfur dioxide (SO2) emissions in the provinces of China (details about variables, study area, and data sources are provided in the next sub-sections). The scatterplot of X and Y and the fitting line show a positive relationship (Fig. 2, left panel). However, when turning to consider the variables expressed as per capita data -\( RX = X / Pop \) and \( RY = Y / Pop \) - a negative relationship can be found (Fig. 2, right panel). The above issues can also be detected by comparing the simple correlation coefficient \( \rho_{XY} = +0.33 \) and the partial correlation coefficient \( \rho_{XY|Pop} = -0.45 \), which changes in sign and size. The previous example shows that the net relationship between several phenomena may be quite different from the gross one, the latter being possibly inflated by latent variables. That implies the need to consider all the potential factors that determine the health status of a society.

2.2. Nomenclature

As far as the dependent variables are concerned, let us denote by \( Cov \) the total number of confirmed cases of 2019-nCoV, and by \( RCoV \) the incidence rate of the infection, namely, the number of cases per 100 thousand inhabitants.

The first set of covariates is as follows:

- \( Pop \) is the population size;
- \( Den \) is the population density, namely, the ratio between population and area in \( \text{km}^2 \);
- \( Grp \) is the gross regional product;
- \( Pr \) stands for the yearly average precipitation;
- \( Th \) indicates the annual average maximum temperature;
- \( SO2 \) is the levels of sulfur dioxide emissions;
- \( Iwg \) stands for the emissions of industrial waste gases.

The second set of covariates includes the variables that depend on human activity and are hence normalized by population:

- \( RGrp = Grp / Pop \) is the per capita gross regional product;
- \( RSO2 = SO2 / Pop \) stands for the per capita emissions of sulfur dioxide;
- \( RIwg = Iwg / Pop \) indicates the emissions of industrial waste gases in per capita values.

https://doi.org/10.1016/j.scitotenv.2020.141028
0048-9697/© 2020 Elsevier B.V. All rights reserved.
2.3. Study area and data sources

This study focuses on 28 mainland Chinese provinces, autonomous regions, and municipalities outside Hubei province. Tibet and Guizhou are excluded due to missing data.

Data concerning the cumulative confirmed cases of the 2019-nCoV as of March 22, 2020 (Fig. 3) - irrespective of whether they resulted in deaths, and regardless of the number of people that have recovered from the virus - is collected from the Coronavirus Resource Center of Johns Hopkins University (Dong et al., 2020).

Demographic and economic variables are derived from the annual publication of the National Bureau of Statistics of China (National Bureau of Statistics of China, 2019). Information about precipitation and temperature is gathered from Current Results,¹ based on data conveyed by the China Meteorological Administration and the World Meteorological Organization.

Data about pollution emissions are taken from the paper "The Pollution state in 31 Provinces and Regions in China" (Yang and Yang, 2011). Although outdated, those values are assumed to be a fair proxy of current emissions in Chinese provinces.

2.4. Analytical models

To study the relationships between dependent and independent variables, we use the spatial autoregressive (SAR) models (Copiello and Grillenzoni, 2017; Elhorst, 2010) with exogenous predictors:

\[
\text{Cov}_i = \alpha + \rho \text{Cov}_{i-1} + \beta_1 \text{Pop}_i + \beta_2 \text{Den}_i + \beta_3 \text{Grp}_i + \beta_4 \text{Pr}_i + \beta_5 \text{Th}_i + \beta_6 \text{SO2}_i + \beta_7 \text{LwG}_i + e_i, \quad e_i \sim \text{IN}(0, \sigma^2_e)
\]

(1)

\[
\text{RCov}_i = \alpha + \rho \text{RCov}_{i-1} + \beta_1 \text{Pop}_i + \beta_2 \text{Grp}_i + \beta_3 \text{Pr}_i + \beta_4 \text{Th}_i + \beta_5 \text{SO2}_i + \beta_6 \text{LwG}_i + u_i, \quad u_i \sim \text{IN}(0, \sigma^2_u)
\]

(2)

where \( i \) is the province index, \( \alpha, \beta_5, \) and \( \rho \) are the coefficients, and \( e_i \) and \( u_i \) are residuals, which are expected to be independent and normally (IN) distributed, with mean zero and constant variance. It is worth noting that the model of Eq. (2) differs from the model of Eq. (1) because the variables which directly depend on human activity are normalized by population in the latter.

In the models of Eqs. (1) and (2), \( \rho \) is the spatial autocorrelation coefficient; hence, \( \text{Cov}_{i-1} \) and \( \text{RCov}_{i-1} \) are spatially lagged dependent variables (i.e., the mean values of \( \text{Cov}_j \) and \( \text{RCov}_j \) in the \( j \) provinces contiguous to the \( i \)th area). These lagged terms aim to identify whether the analyzed phenomenon has a spatial pattern accordingly to Tobler's (1970) first law of geography, namely, that “everything is related to everything else, but near things are more related than distant things” (p. 236).

In order to satisfy the assumption of homoscedasticity (i.e., \( \sigma^2 \) independent of \( i \), all variables are transformed with natural logarithms (ln).

The core of the analysis is represented by statistical significance and sign (+ or -) of the estimated coefficients \( \beta_i \). In particular, the differences in the \( \beta_i \) of in the models of Eqs. (1)–(2) is a symptom of spurious correlation between epidemic and environmental variables.

3. Results and discussion

The estimates of the models of Eqs. (1) and (2) are provided in Tables 1 and 2 (see also Figs. 4 and 5). In general, they are satisfactory as they fulfill the standard assumptions of regression, namely, normal residuals, absence of outliers and multicollinearity, and good fitting. Specifically, the hypothesis \( H_0: e_i, u_i \sim \text{IN}(0, \sigma^2) \) is accepted with low Chi^2(2) statistics: 2.8637 (p-value 0.2389) for Cov, and 0.7412 (p-value 0.6903) for RCov. The explanatory variables are not affected by multicollinearity according to the low Variance Inflation Factors (VIFs ≤2.5, suggested in Allison, 1999). The adjusted R^2 are 0.7124 for Cov, and 0.4906 for RCov.

The coefficients of the spatially lagged terms are not significant, meaning the absence of spatial correlation in the data. However, that may depend on the high level of spatial aggregation of provincial data, which involves suitable local policies to control the epidemic. For example, movement restrictions should be better adopted at the national level, at least, provided the national borders are not porous.

As regards the analysis of the explanatory variables, in the model with Cov (the absolute number of confirmed cases), the average maximum temperature \( \text{Th} \) plays a significant role. Apart from indirect effects - namely, the higher is the temperature, the higher is the level of social interactions, and so the spread of the infection - the positive coefficient of \( \text{Th} \) stimulates other interpretations. Assuming that the novel coronavirus was already circulating before December 2019, it could imply that the recent global outbreak is also related to the mild weather conditions experienced in February 2020 (Masters, 2020). That contrasts with the expectation that the epidemic will spread less easily and move slowly during spring and summer as temperatures get warmer, as also suggested in other articles published in the journal Science of the Total Environment (Ma et al., 2020; Xie and Zhu, 2020). However, it has to be
considered that the incidence rate of the infection - adjusted for population density and other factors - has been found to be inversely associated with warmer and drier weather conditions (Byass, 2020).

Another significant covariate of the overall confirmed cases is the gross regional product $\text{Grp}$, which takes on a positive sign. Incidentally, that predictor is significantly correlated with some of the variables representing air pollution ($\text{SO}_2$: $\rho = 0.5516$, $p$-value $0.0023$; $\text{Iwg}$: $\rho = 0.5412$, $p$-value $0.0029$). Apparently, this finding confirms the association found in the study authored by Zhu et al. (2020). Nevertheless, it is a trivial result. It has to be expected that the

| Table 1 | Results of the estimation of the model of Eq. (1) for the overall confirmed cases of 2019-nCoV. |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| Predictor       | Coefficient     | Std. err.       | t-Stat           | p-Value         | VIF             |
| const           | -5.997          | 1.256           | -4.773***        | 0.0001          | -               |
| $\text{Grp}$    | 0.889           | 0.110           | 8.078***         | 0.0000          | 1.085           |
| $\text{Th}$     | 0.966           | 0.274           | 3.532***         | 0.0016          | 1.085           |

$\text{Cov}$: total number of confirmed cases of 2019-nCoV. $\text{Grp}$: gross regional product. $\text{Th}$: annual average maximum temperature. $^1$ Significance levels: * 0.1; ** 0.05; *** 0.01.

| Table 2 | Results of the estimation of the model of Eq. (2) for the incidence rate of 2019-nCoV. |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| Predictor       | Coefficient     | Std. err.       | t-Stat           | p-Value         | VIF             |
| const           | -1.241          | 0.367           | -3.378***        | 0.0024          | -               |
| $\text{Den}$    | 0.286           | 0.047           | 6.056***         | 0.0000          | 1.022           |
| $\text{Rhwg}$   | -0.528          | 0.180           | -2.941***        | 0.0069          | 1.022           |

$\text{RCov}$: number of cases of 2019-nCoV per 100 thousand inhabitants. $\text{Den}$: population density. $\text{Rhwg}$: per capita emissions of industrial waste gases. $^1$ Significance levels: * 0.1; ** 0.05; *** 0.01.
overall confirmed cases are higher in the most populated areas, which are usually also the most industrialized and wealthy, and, as a consequence, the most polluted ones.

The problem is much more evident when turning to the analysis of the predictors of the incidence rate $RCov$. Population density is a significant driver of the number of cases per 100,000 population. That is in keeping with earlier literature (Amuakwa-Mensah et al., 2017), as well as with recent studies focusing on how population size and population density affect both the current and future spread of COVID-19 disease (Jahangiri et al., 2020; Rocklöv and Sjödin, 2020; Zhang et al., 2020). That might explain why, to date, the epidemic has hit so hard several highly densely populated areas around the world: Lombardy region in Italy, North Rhine-Westphalia in Germany, Madrid metropolitan area as far as Spain is concerned, New York in the United States, San Paulo in Brazil, and so forth.

That is actually the issue with the finding presented by Zhu et al. (2020), namely, that the authors missed normalizing by population the number of novel coronavirus cases before testing the relationship with air pollution and other covariates. The authors state that their generalized additive model also includes “city fixed effects ... to control for time-invariant city characteristics such as population size and density” (p. 2). Unfortunately, the results are not reported in full detail, so it is unknown the role played by those fixed effects, as well as how the same fixed effects interact with the variables measuring the pollutants. Hence, there remains an open question: would the COVID-pollution relationship be confirmed using the incidence rate, instead of the number of confirmed cases, as the dependent variable?

Furthermore, in the model of Eq. (2), the level of emissions of industrial waste gas $Iwg$ is another significant predictor of the number of cases per 100,000 population. Nevertheless, it takes on a negative sign, which leaves room for doubt about the hypothesis that air pollution has actually played a role in the spread of 2019-nCoV.

**Declaration of competing interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**References**

Adda, J., 2016. Economic activity and the spread of viral diseases: evidence from high frequency data. Q. J. Econ. 131, 891–941. https://doi.org/10.1093/qje/qjw005.

Allison, P.D., 1999. Multiple Regression: A Primer. Pine Forge Press, Thousand Oaks.

Amuakwa-Mensah, F., Marbuah, G., Mubanga, M., 2017. Climate variability and infectious diseases nexus: evidence from Sweden. Infect. Dis. Model. 2, 203–217. https://doi.org/10.1016/j.idm.2017.03.003.

Becker, S., Soukup, J.M., 1999. Exposure to urban air particulates alters the macrophage-mediated inflammatory response to respiratory viral infection. J. Toxicol. Environ. Heal. A 57, 445–457. https://doi.org/10.1080/00984109915753539.

Byass, P., 2020. Eco-epidemiological assessment of the COVID-19 epidemic in China, January–February 2020. Glob. Health Action 13. https://doi.org/10.1080/16549716.2020.1760490.

Cheng, Z.J., Shan, J., 2020. 2019 novel coronavirus: where we are and what we know. Infection 48, 155–163. https://doi.org/10.1007/s15010-020-01401-y.

Ciencewicki, J., Jaspers, I., 2007. Air pollution and respiratory viral infection. Inhal. Toxicol. 19, 1135–1146. https://doi.org/10.1080/08958370701685434.

Coccia, M., 2020. Factors determining the diffusion of COVID-19 and suggested strategy to prevent future accelerated viral infectivity similar to COVID. Sci. Total Environ. 729, 138474. https://doi.org/10.1016/j.scitotenv.2020.138474.
Conticini, E., Fediani, B., Caro, D., 2020. Can atmospheric pollution be considered a cofactor in extremely high level of SARS-CoV-2 lethality in northern Italy? Environ. Pollut. 261, 114465. https://doi.org/10.1016/j.envpol.2020.114465.

Copiello, S., Grillenzoni, C., 2017. Is the cold the only reason why we heat our homes? Empirical evidence from spatial series data. Appl. Energy 193, 491–506. https://doi.org/10.1016/j.apenergy.2017.02.011.

Cui, Y., Zhang, Z.-F., Froines, J., Zhao, J., Wang, H., Yu, S.-Z., Detels, R., 2003. Air pollution and case fatality of SARS in the People’s Republic of China: an ecologic study. Environ. Health 2, 1–5. https://doi.org/10.1186/1476-069x-2-15.

Dong, E., Du, H., Gardner, L., 2020. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis. 20(5), 5–20. https://doi.org/10.1016/S1473-3099(20)30120-1.

Elhorst, J.P., 2010. Applied spatial econometrics: raising the bar. Spat. Econ. Anal. 5, 9–83. https://doi.org/10.1080/17421770903541772.

Horne, R.D., Joy, E.A., Hofmann, M.G., Gesteland, P.H., Cannon, J.B., Leffler, J.S., Blagey, D.P., Korgenski, E.K., Torosyan, N., Hansen, G.J., Karchtner, D., Pope, C.A., 2018. Short-term elevation of fine particulate matter air pollution and acute lower respiratory infection. Am. J. Respir. Crit. Care Med. 198, 759–766. https://doi.org/10.1164/rccm.201709–1883OC.

Iqbal, N., Fareed, Z., Shahzad, F., He, X., Shahzad, U., Lina, M., 2020. The nexus between COVID-19, temperature and exchange rate in Wuhan city: new findings from partial and multiple wavelet coherence. Sci. Total Environ. 729, 138916. https://doi.org/10.1016/j.scitotenv.2020.138916.

Jahangiri, Mehdi, Jahangiri, Milad, Najafgholipour, M., 2020. The sensitivity and specificity analyses of ambient temperature and population size on the transmission rate of the novel coronavirus (COVID-19) in different provinces of Iran. Sci. Total Environ. 728, 138872. https://doi.org/10.1016/j.scitotenv.2020.138872.

Jia, J.S., Lu, X., Yuan, Y., Xu, G., Jia, J., Christakis, N.A., 2020. Population flow drives spatio-temporal distribution of COVID-19 in China. Nature 582, 389–394. https://doi.org/10.1038/s41586-020-2284-y.

Ma, Y., Zhao, Y., Liu, J., He, X., Wang, B., Fu, S., Yan, J., Niu, J., Zhou, J., Luo, B., 2020. Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China. Sci. Total Environ. 724, 138226. https://doi.org/10.1016/j.scitotenv.2020.138226.

Masters, J., 2020. February 2020: earth’s 2nd warmest February and 3rd warmest month on record. Sci. Am. https://blogs.scientificamerican.com/eye-of-the-storm/beginning-of-earth-2nd-warmest-february-and-3rd-warmest-month-on-record/ (last accessed 16.07.2020).

Mehta, S., Shin, H., Burnett, R., North, T., Cohen, A.J., 2013. Ambient particulate air pollution and acute lower respiratory infections: a systematic review and implications for estimating the global burden of disease. Air Qual. Atmos. Health 6, 69–83. https://doi.org/10.1007/s11869-011-0146-3.

National Bureau of Statistics of China, 2019. China Statistical Yearbook. China Statistics Press, Beijing.

Rocklov, J., Sjödin, H., 2020. High population densities catalyse the spread of COVID-19. J. Travel Med. 27, 1–2. https://doi.org/10.1093/jtm/taaa038.

Sarmadi, M., Marufi, N., Kazemi Moghaddam, V., 2020. Association of COVID-19 global distribution and environmental and demographic factors: an updated three-month study. Environ. Res. 188, 109748. https://doi.org/10.1016/j.envres.2020.109748.

Sobral, M.F.F., Duarte, G.B., da Penha Sobral, A.L.G., Marinho, M.L.M., de Souza Melo, A., 2020. Association between climate variables and global transmission of SARS-CoV-2. Sci. Total Environ. 729, 138997. https://doi.org/10.1016/j.scitotenv.2020.138997.

Tobler, A.W.R., 1970. A computer movie simulating urban growth in the Detroit region. Econ. Geogr. 46, 234–240. https://doi.org/10.2307/143141.

Xie, J., Zhu, Y., 2020. Association between ambient temperature and COVID-19 infection in 122 cities from China. Sci. Total Environ. 724, 138201. https://doi.org/10.1016/j.scitotenv.2020.138201.

Xu, Q., Li, X., Wang, S., Wang, C., Huang, F., Gao, Q., Wu, L., Tao, L., Guo, J., Wang, W., Guo, X., 2016. Fine particulate air pollution and hospital emergency room visits for respiratory disease in urban areas in Beijing, China, in 2013. PLoS One 11, e0153099. https://doi.org/10.1371/journal.pone.0153099.

Yang, J.Q., Yang, S., 2011. The population size in 31 provinces and regions in China. Procedia Environ. Sci. 11, 355–357. https://doi.org/10.1016/j.proenv.2011.12.057.

Ye, Q., Fu, J., Mao, J., Shang, S., 2016. Haze is a risk factor contributing to the rapid spread of respiratory syncytial virus in children. Environ. Sci. Pollut. Res. 23, 20178–20185. https://doi.org/10.1007/s11356-016-7228-6.

Zhang, X., Liu, H., Tang, H., Zhang, M., Yuan, X., Shen, X., 2020. The Effect of Population Size for Pathogen Transmission on Prediction of COVID-19 Pandemic Spread. pp. 1–30.

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.

Sergio Copiello* Carlo Grillenzoni
IUAV University of Venice, Dorsoduro 2206, 30123 Venice, Italy
*Corresponding author.
E-mail address: sergio.copiello@iuav.it.

23 June 2020