A Multiplicity of Environmental, Economic and Social Factor Analyses to Understand COVID-19 Diffusion

Juan Qiu (✉ qiujuan@apm.ac.cn)
Institute of Geodesy and Geophysics. Chinese Academy of Sciences  https://orcid.org/0000-0001-9763-2706

Rendong Li
Chinese Academy of Sciences Innovation Academy for Precision Measurement Science and Technology

Dongfeng Han
Chinese Academy of Sciences Innovation Academy for Precision Measurement Science and Technology

Qihui Shao
Chinese Academy of Sciences Innovation Academy for Precision Measurement Science and Technology

Yifei Han
Chinese Academy of Sciences Innovation Academy for Precision Measurement Science and Technology

Xiyue Luo
Hubei University

Yanlin Wu
Hubei University

Research

Keywords: COVID-19, Human-environment system, Determinant, Public health control measures, Mixed-effects mixed-distribution model

Posted Date: July 21st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-717943/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Version of Record: A version of this preprint was published at One Health on December 1st, 2021. See the published version at https://doi.org/10.1016/j.onehlt.2021.100335.
A multiplicity of environmental, economic and social factor analyses to understand COVID-19 diffusion

Juan Qiu1*, Rendong Li1, Dongfeng Han1,2, Qihui Shao1,2, Yifei Han1,2, Xiyue Luo3 and Yanlin Wu3

1 Key Laboratory of Monitoring and Estimate for Environment and Disaster of Hubei Province, Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences, Wuhan 430077, China
2 University of Chinese Academy of Sciences, Beijing 100049, China
3 Faculty of Resources and Environmental Science, Hubei University, Wuhan 430062, China

*Correspondence: qiujuan@apm.ac.cn

ABSTRACT

Background: Research on the impact of the environment on COVID-19 lacks a full-comprehensive perspective, and neglecting the multiplicity of the human-environment system can lead to misleading conclusions. We attempted to reveal all environmental-to-human and human-to-human determinants that influence the transmission of COVID-19.

Methods: We estimated the daily case incidence ratios (CIR) of COVID-19 for prefectures across mainland China, and used a mixed-effects mixed-distribution model to study the association between the CIR and 114 considered factors related to climate, atmospheric environmental quality, terrain, population, economic, human mobility as well as public health control measures (PHCMs).

Results: CO, O₃, PM₁₀ and PM₂.₅ were positively linked with CIR, but the effect of NO₂ was negative. The temperature had no significant association with CIR, and the daily minimum humidity was a significant negatively predictor. National emergency response level was negatively associated with CIR until with a lag of 15 days. Higher accumulated destination migration scale flow from the epicenter and lower distance to the epicenter (DisWH) were associated with a higher CIR, however, the interaction between DisWH and the time was positive.

In China, the more economically developed and more densely populated cities have a higher probability of CIR occurrence, but they may not have a higher CIR intensity.
Conclusion: The COVID-19 pandemic’s diffusion patterns are caused by a multiplicity of environmental, economic, social factors as well as public health control measures PHCMs. First, multiple pollutants carried simultaneously on particulate matter act in a synergistic way to affect COVID-19 transmission. Second, the temperature has a limited impact on the spread of the epidemic. Third, PHCMs must last for at least 15 days or longer before the effect has been apparent. Fourth, preventing the introduction of the epidemic between cities seemed to be more complex and difficult than controlling the growth of the epidemic within the city. China’s relatively cautious response to the epidemic could provide lessons for countries that are still experiencing a health emergency and help to prevent future pandemics similar to COVID-19.

Keywords: COVID-19, Human-environment system, Determinant, Public health control measures, Mixed-effects mixed-distribution model

Introduction

Globally, as of 7 July 2021, there have been 184,324,026 confirmed cases of novel coronavirus disease (COVID-19), including 3,992,680 deaths [1]. A better understanding of the effects of comprehensive natural and human environmental factors on COVID-19 transmission could contribute to finding solutions for its monitoring and treatment. In addition, environmental changes, such as climate change, land use changes, urbanization, biodiversity loss, and invasive species, may increase the risk of emerging infectious diseases (EIDs) [2-4], and the environmentally related experience gained from COVID-19 will also provide lessons for future prevention and control of EIDs.

Regarding this ongoing major health crisis, the natural environment that researchers focus on is the atmospheric environment, specifically airborne particulate matter such as PM$_{2.5}$ and PM$_{10}$, meteorological variables such as temperature and relative humidity and air pollutants such as SO$_2$ and NO$_2$ [5-13]. In addition, the human environment that researchers focus on is socioeconomic, such as GDP, demographic variables, such as population density, and human activity patterns, such as human mobility and control measures [14-20]. The pandemic is actually a very complex phenomenon, and its diffusion patterns are typically caused by a multiplicity of environmental, economic and social factors [21]; hence, full-comprehensive variables and systematic analytical procedures should be considered in the analysis to describe any possible correlation in a rigorous
way [13, 22-24]. However, current studies did not consider that complex outcomes may be due to disciplinary specialties increasing attitudes of scholars concentrating on specific factors, neglecting this multiplicity during a pandemic crisis that can present substantial risks of bias and lead to misleading conclusions [23, 25].

Public health control measures (PHCMs) such as closing schools, closing entertainment venues and suspending intracity public transport (bus and subway), banning public gatherings and intercity travel restrictions have been proven to have a positive effect on the control of COVID-19 [9, 14-16, 21, 26-28]. However, almost all of these control measures are simulated by an infectious disease dynamics model, and due to the lack of detailed quantitative data on prevention and control measures, the control measures are hardly statistically analyzed in the early stages of the epidemic.

This novel coronavirus was identified in December 2019 in Wuhan City of China, and since 23 January, thirty provinces, municipalities and autonomous regions in China sequentially activated the Level-I alert of public health incidents, the highest level of emergency public health alerts and responses in the nation's public health management system [29]. After reaching a peak on February 13, 2020, the number of newly confirmed cases gradually decreased, and after April 3, the number of new local cases remained in single digits or 0, followed by small-scale outbreaks (the number of new cases per day is less than 300) in several cities since June 11 [30, 31] that appear to be related to imported cases or seafood [32]. Therefore, the period from the first confirmed case in Wuhan to June 11, 2020, in China was a perfect natural example of the epidemic evolution that can reflect the impact of the integrated human-environment system on the development of the epidemic.

Therefore, this study will conduct an evaluation in China on developing a comprehensive vision of COVID-19 contagion that considers a multiplicity of transmission channels, including the natural environment (environmental-to-human) and human channels (human-to-human). At the same time, the delay effects of control measures, population movement and atmospheric environment are also discussed.

**Materials and methods**

**Data collection and preparation**

In the study, city-level data were collected among 366 cities of 31 provinces in mainland
China (Fig. 1) between Jan 10 and Jun 11, 2020. Daily data on the number of newly confirmed COVID-19 cases were obtained from the Outbreak Notification of the National Health Commission of the People’s Republic of China (http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml) and provincial or municipal health commissions. The accumulated confirmed COVID-19 cases per 100,000 persons at the city level until June 11, 2020 are shown in Fig. 1. The number of newly confirmed cases of novel coronavirus pneumonia (NCP) on 12 February released by Hubei Province increased significantly because it included 13,332 cases confirmed by clinical diagnosis instead of previously relying on nucleic acid testing for the confirmation of cases [33]. To reflect the true daily new confirmed cases, we evenly distributed the increment of cases on February 12 to each previous day from January 10th to February 12th (see Supplementary Figure 1).

The association between an independent factor and disease incidence counts may fail to infer the association with disease transmission [22], so we transformed the incidence data to case incidence ratios ($CIR$, cases per 10 million persons) reflected by both the number of COVID-19 cases and population base and checked the association between CIR and external factors (Table 1) thereafter.

The time variable ($Day$) from January 10, 2020 to June 11, 2020 was coded from 0 to 154. Considering that the change in CIR with time is a curve process of first increasing and then decreasing, the quadratic growth model is used to add a square term of the time variable, namely, $Day^2 = Day \times Day$.

The daily climate data, including daily minimum temperature ($MinT$), maximum temperature ($MaxT$), mean temperature ($MeanT$), relative humidity ($Rh$) and minimum relative humidity ($MinRh$), were downloaded from the China Meteorological Data Service Center (http://data.cma.cn).

Atmospheric environmental quality data, including daily CO, NO$_2$, O$_3$, PM$_{2.5}$, PM$_{10}$, SO$_2$ and AQI (air quality index), were downloaded from the Data Center of the Ministry of Ecology and Environment of the People’s Republic of China (https://datacenter.mee.gov.cn/). Linear interpolation was used to replace missing values. Meanwhile, consider the delay effect of atmospheric environmental indicators. For example, $CO_1$ is defined as CO delayed by one day, etc., $CO_9$ is defined as CO delayed by 9 days. Other atmospheric environmental indicators are
similar.

The terrain data ASTER GDEM 30M were downloaded from the Geospatial Data Cloud, Computer Network Information Center, Chinese Academy of Sciences (http://www.gscloud.cn). The mean DEM (MeanDEM) of each city was calculated.

Among the human factors, the household population (Pop) and gross domestic product (GDP) in 2019 were collected from statistical yearbooks of each province of China. At the same time, considering the interaction effect of GDP and time, it is named \(Day\_GDP = Day \times GDP\).

The population density (PD) in 2019 was downloaded from the Ministry of Housing and Urban-Rural Development of China (http://www.mohurd.gov.cn/xytj/index.html) and refers to the density of the population in an urban area.

The destination proportion in population flow from Wuhan (WH) and migration scale (MS) from January 1 to January 23 (Wuhan lockdown), 2020, were downloaded from Baidu (http://qianxi.baidu.com/), and the destination migration scale flow from Wuhan (Popmob) was calculated by multiplying WH by MS. Meanwhile, considering the delay effect of Popmob, Popmob1 is defined as a delay of one day, etc., Popmob9 is defined as a delay of 9 days. The cumulative Popmob (Popmobsum) from January 1st to the current date was also calculated.

The distance of each city from Wuhan (DisWH) was calculated. At the same time, considering the interaction effect of DisWH and time, it is named \(Day\_DisWH = Day \times DisWH\).

According to the national emergency plan for public health emergencies (http://www.gov.cn/yjgl/2006-02/26/content_211654.htm), China's public health alert system is categorized into four levels in terms of the nature of the incidents, extent of harm and scope: Level-I (extremely significant), Level-II (significant), Level-III (major) and Level-IV (normal). The description of each emergency response level and the corresponding specific measures in response to this epidemic are shown in the Supplementary Table 1. The level of the adopted control measures (Reslevel) at different times in 366 cities were collected from National Health Commission of the People’s Republic of China (http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml) and provincial or municipal health commissions. We recoded no response, Level-IV, Level-III, Level-II and Level-I as 0, 1, 2, 3 and 4, where the larger the value of Reslevel, the stricter the measures. Meanwhile, considering the delay effect of Reslevel, Reslevel1 is defined as a delay of one day, etc., Reslevel20 is defined as a
delay of 20 days.

Models

The CIRs were extremely nonnormally distributed, exhibiting a large clump of values at zero (51776/56364, 91.8%) and skewed nonzero values (see Supplementary Table 2, Figure 2 and Figure 3). To address these semicontinuous outcome measures, we used a mixed-effects mixed-distribution model (also called a multilevel mixed distribution model) with correlated random effects for repeated measures data with clumping at zero and highly skewed [34, 35]. The model contains components to model the occurrence probability of a nonzero value (the ‘occurrence model’, based on logistic regression using all city-day CIR data) and the probability distribution of nonzero values (the ‘CIR intensity’ model, based on lognormal regression using city-day records where CIR > zero), allowing for repeated measurements using random effects and allowing for correlation between the two components. The correlation allows cities with higher rates of occurrence (specifies whether there are any new confirmed cases) to also have higher (or lower) mean nonzero responses (specifies the CIR nonzero value).

An SAS Macro, MIXCORR [35] was used for this analysis. The approach is based on maximum likelihood estimation for estimating the effect of explanatory variables on the probability of nonzero values, the mean of nonzero values, and the overall mean amount. The model forms were:

**Occurrence model (a logistic regression model):**

$$\text{logit}(p_{ij}) = \beta_{10} + \beta_{11} \text{Day}_{ij} + \beta_{111} \text{Day}_{ij}^2 + \sum_{k=1}^{K} \beta_{1k} X_{1kj} + \mu_{ij}$$

$$R_{ij} \begin{cases} 0, & \text{if } \text{CIR}_{ij} = 0 \\ 1, & \text{if } \text{CIR}_{ij} > 0 \end{cases}$$

where \(p_{ij}\) is the probability of having nonzero CIR values (\(R_{ij} = 1\), \(R_{ij}\) represents the occurrence variable), \(\text{CIR}_{ij}\) is city \(i\)’s CIR on Day \(j\), and \(X_{1kj}\) is a vector of covariates that explain \(\Pr (R_{ij} = 1)\). Since the change in CIR with time is nonlinear, the polynomial curve development model needs to be considered. In this study, a quadratic term of Day, namely, \(\text{Day}^2 (\text{Day}_{ij}^2 = \text{Day}_{ij} \ast \text{Day}_{ij})\), is constructed in the model. Regression coefficients \(\beta_{10}, \beta_{11}, \beta_{111}\), and \(\beta_{1k}\) are the intercept, fixed time (Day) effect, fixed time (Day^2) effect and fixed effects of covariates on the
log-odds of $R_{ij} = 1$, respectively, and $\mu_{ij}$ is the random effect of individuals on the log-odds. In these multilevel modeling terms, $(\beta_{10} + \mu_{ij})$ is the random intercept that allows the probabilities of having a nonzero CIR to vary across cities.

CIR intensity model (a linear regression model):

$$\log(S_{ij}) = \beta_{20} + \beta_{21} \text{Day}_{ij} + \beta_{211} \text{Day}_{ij}^2 + \sum_{k=1}^{K} \beta_{2j} X_{2kj} + \mu_{2j} + e_{ij}$$

where $S_{ij}$ is the intensity variable defined $S_{ij} = (\text{CIR}_{ij} | R_{ij} = 1)$, $X_{2kj}$ is a vector of covariates for intensity that explain nonzero CIR$_{ij}$, $e_{ij}$ is the level-1 residuals, and $\mu_{2j}$ is the random effect on the initial level of intensity. The random intercept ($\beta_{20} + \mu_{2j}$) in the CIR intensity model accounts for the heterogeneity of the mean nonzero CIR among cities. The two random effects $\mu_{1j}$ and $\mu_{2j}$ are assumed to be jointly normally distributed.

Initially, there were 112 variables (predictors) and 56,364 observations in this study. Faced with such high-dimensional independent variables, the variables were selected using the Lasso [36], which has been proven to be useful and feasible when the number of observations is much larger than the number of predictors [37, 38]. Then, the selected variables were incorporated into the mixed-effects mixed-distribution model. The data preprocessing and descriptive statistical analysis were performed with SPSS software version 25 (IBM), variables were selected in Stata 16 (StataCorp), and the modeling analysis was conducted using SAS 9.4 (SAS). Spatial analysis and map creation were performed in ArcGIS 10.7 (ESRI).

Results

Variable selection

By the Lasso method, 57 (Table 2) of the original 112 variables (Table 1) were selected. Among these 57 variables, except Pop, PD, GDP, DisWH, MeanDEM, MinT and MinRh, all other variables considered the delay effect. In Lasso or OLS parameter estimation, the slopes of variables Popmob0, Popmob1, Popmob2, Popmob3, Popmob4 and Popmob5 were negative, and the slopes of variables Popmob5, Popmob5 and Popmobsum were positive. Considering that the incubation period of COVID-19 ranges from 2 to 14 days or even longer [39, 40], the number of new cases on a certain day is likely to be affected by the cumulative number of inflows from Wuhan before that day, rather than the number of inflows on that day or a day before. Therefore, Popmobsum was selected for the next modeling. For national public health emergency responses,
the slopes of variables Reslevel2, Reslevel3, Reslevel5, Reslevel6, Reslevel8 and Reslevel9 were positive, and the slopes of variables Reslevel15, Reslevel17 and Reslevel20 were negative. The positive effects of the control measures on the epidemic seem to be reflected after 15 days, so Reslevel15 was selected for the next modeling. The 23 selected variables that reflected the atmospheric environmental quality were almost time-continuous delay indicators, such as NO₂; ranging from no delay (NO₂) to nine-day delay (NO₂_9). The variables with the longest delay time were selected to enter the next modeling, such as CO_9, SO₂_9, NO₂_9, O₃_9, PM10_9 and PM25_8. In summary, there were a total of 15 variables mentioned above plus 2 interaction effect variables Day_GDP and Day_DisWH for the mixed-effects mixed-distribution modeling.

**Modeling analysis**

The SAS macro outputted fitting statistics AIC and -2LL in the case of uncorrelated and correlated random effects, respectively (Table 3). The difference of -2LL between the two models was 54.33, and the corresponding chi-square test P value was less than 0.0001, indicating that the model with correlated random effects fit better. Therefore, the parameter estimation of the model with correlated random effects is listed in Table 3, which examines the multiple comprehensive effects of natural and human environments, including 17 variables that reflect climate, terrain, atmospheric environmental quality, population, economic, human mobility and PHCMs on the COVID-19 CIR at the city level.

The linear and quadratic effects of the time variables (Day and Day2) were statistically significant both on the probability of CIR occurrence ($\hat{\beta}_{11} = 0.1765, p < 0.0001; \hat{\beta}_{111} = -0.0030, p < 0.0001$) and on the proportion of CIR ($\hat{\beta}_{21} = -0.0115, p = 0.0018; \hat{\beta}_{211} = -0.0003, p < 0.0001$), indicating that the probability of CIR occurrence first increased and then decreased, and the inflection point was on February 8, 2020 (Day=29.4). However, the CIR intensity has continued to decrease since confirmed cases have been reported in a certain area (Day > 0).

Fifteen variables and interaction effects related to climate, atmospheric environmental quality, terrain, population, economic, human mobility and PHCMs had a statistically significant effect on the probability of CIR occurrence, except MinT ($p = 0.1106$) and S02_9 ($p = 0.7423$) at the 95% confidence level (Table 3). Specifically, the CIR occurrence was positively associated with Pop, PD, GDP, Popmobsum, and MinRh, as well as the atmospheric environmental quality index CO_9, O₃_9, PM10_9 and PM25_8 and negatively associated with MeanDEM, DisWH,
Reslevel15 and NO2_9. Furthermore, the interaction of GDP, DisWH and Day (Day_GDP and Day_DisWH, respectively) was statistically significantly positively correlated with CIR occurrence. However, eleven variables and interaction effects had a statistically significant effect on CIR intensity, except PD, MeanDEM, MinRh and PM25_8, which were statistically significant for CIR occurrence at the 95% confidence level. Higher GDP, Popmobsum, O3_9, PM10_9 and Day_DisWH were associated with a higher CIR intensity, but lower Pop, DisWH, Reslevel15, MinRh, NO2_9 and Day_GDP were associated with a higher CIR intensity. Mean fixed (background) determinant distributions and determinants that changed over time on CIR incidence or intensity detected by the mixed-effects mixed-distribution model are shown in Fig.2 and Fig. 3. In general, PM2.5 and CO showed a downward trend, PM10 fluctuated, NO2 first decreased and then increased slightly after the national emergency response intensity gradually decreased, O3 showed an upward trend, and daily minimum temperature (MinT) first decreased in a fluctuating way and then increased slightly after May (Fig. 3). The popmobsum continued to increase before the Wuhan lockdown and remained unchanged after the Wuhan lockdown. The reslevel rapidly rose to the highest level on January 23 and gradually decreased after February 21 (Fig. 3).

The significant random effects variance for both the CIR occurrence and CIR intensity (\( \hat{\sigma}_1^2 = 0.9592, p < 0.0001; \hat{\sigma}_2^2 = 0.4287, p < 0.0001 \)) showed that there were great individual differences between cities in both the probability of CIR occurrence and the CIR intensity. The significant positive correlation between CIR occurrence and intensity random effects (\( \hat{\rho}_{\sigma_1\sigma_2} = 0.2819, p <0.0001 \)) indicated that, on average, cities with a higher likelihood of CIR occurrence tended to report a higher mean amount of CIR.

**Discussion**

In this study, a mixed-effects, mixed-distribution model for longitudinal data identified a subset of environmental-to-human and human-to-human determinants of COVID-19 occurrence and intensity and their changes over time as the pandemic progressed. To the best of our knowledge, our study is the first to more comprehensively examine multiple longitudinal data sets to understand COVID-19 occurrence and intensity risk factors among cities across China. We found that many factors related to terrain, climate, atmospheric environmental quality, population, economy, human mobility and PHCMs synergistically affect the spread of the
epidemic. However, the determinants that significantly affect the occurrence probability are broader than the occurrence intensity, which seems to imply that preventing and controlling the introduction of the epidemic between cities seems to be more difficult and complex than controlling the growth of the epidemic within the city.

Our results demonstrated that higher population size and density were associated with an increased risk of COVID-19 occurrence but with a lower CIR intensity, and the effect of population density on CIR intensity was not statistically significant. Other studies for COVID-19 have proven that population density affects the number of COVID-19 daily cases [17, 41] but is not associated with accumulated COVID-19 cases [42]. The difference in those results was due mainly to the different dependent variables reflecting the COVID-19 epidemic index. In addition, our population density refers to the density of the population in an urban area rather than the whole city administrative area, and the former can better reflect the degree of natural contact between people.

Cities with higher GDP had more CIR occurrence and intensity (Table 3). Our results parallel prior research, which found that the number of COVID-19 cases and deaths is higher in high-income countries [18]. We further considered whether and how the impact of GDP on the epidemic changes over time. The interaction effect among the time variable and GDP indicated that over time, cities with high GDP had a greater increase in the probability of COVID-19 occurrence but a decreased growth rate of CIR intensity. Our results seem to suggest that cities with a higher population size and density have greater intercity population mobility and therefore a higher risk of initial case introduction. To conclude, in China, the more economically developed and more densely populated cities may have a higher probability of occurrence due to the greater intensity of human and logistics flow, but they may not have a higher occurrence intensity due to stronger prevention and control awareness [43] and more effective prevention and control strategies accumulated over time.

We generally found that a lower mean DEM increased the probability of COVID-19 occurrence and CIR intensity, but the result in CIR intensity was not statistically significant. The terrain of China, low in the east and high in the west (Fig. 2), affects population distribution and transportation accessibility, specifically a higher population (Fig. 2) and more convenient transportation in the east [44-46].
Farther away from Wuhan was a protective factor for COVID-19 transmission. This finding aligns with other distances to Wuhan and COVID-19 spread research [47, 48]. However, the positive interaction effect among the time variable and the distance of each city from Wuhan indicated that over time, this effect gradually diminished. Unsurprisingly, the cumulative destination migration scale flow from Wuhan was positively connected with both COVID-19 occurrence and CIR intensity. The relatively high cumulative destination migration scale flow from Wuhan from January 1 to January 23 (Wuhan lockdown), 2020, is distributed mainly in the surrounding cities of Wuhan, especially in the cities of Hubei Province (Fig. 2), which aligns with the high value of accumulated confirmed COVID-19 cases per 100,000 persons distributed mainly in Hubei Province (Fig. 1). Wuhan's geographical location and transportation hub in China has enabled the virus to perforate throughout China [49].

The positive correlation between CIR and the level of PHCMs with a delay of 2 to 9 days indicated that more stringent prevention and control measures were being taken in the more severe epidemic cities in the early stages of the epidemic, while the negative correlation with CIR after a delay of 15 days indicated that the control effect (both the probability of CIR occurrence and CIR intensity declined) did not appear until 15 days after the measures were taken. Other studies have reached similar conclusions; for example, non-pharmaceutical interventions put in place by governments may not have had a significant impact on the initial growth of COVID-19 [50], the impact of non-pharmaceutical interventions on cumulative confirmed cases per million population became visible with a time lag of approximately 5 weeks in Sweden [51], and the effect of introducing on reproduction number was delayed by 1-3 weeks at the country level [52].

Surprisingly, temperature, including the daily minimum temperature, is not significantly related to COVID-19. Most studies suggest that a negative correlation exists between temperature and the number of COVID-19 incidence and severity [53], while there have been studies that support the absence of any correlation or even a positive one [54]. Our findings used data from China to show that the effect of temperature on COVID-19 may not be negligible, yet it was not detected due to the stronger effect of covariates related to the human environment, such as economic, population and public control measures.

In contrast, the daily minimum humidity is more likely to affect the epidemic. Specifically, a higher daily minimum relative humidity was associated with an increased risk of CIR occurrence.
but with a lower CIR intensity. In previous studies, relative humidity was inversely associated with increased cases [55] and mortality rate [56], and a positive correlation or negative correlation was found between relative humidity and confirmed cases in different study areas in Italy [57], while the association between absolute humidity and epidemic growth was no longer significant [58]. The differences in these results may be due to different independent variables reflecting humidity, dependent variables reflecting the COVID-19 epidemic, research period, or research area and scale. Although the effect of humidity on COVID-19 transmission remains obscure, our detected determinant, the daily minimum relative humidity rather than relative humidity, may provide a new perspective for further research.

All available research thus far shows a positive correlation between air pollution (PM$_{2.5}$) and COVID-19 [59]; however, whether multiple pollutants carried simultaneously on particulate matter act in an additive, synergistic way to increase the severity of COVID-19-like diseases is unknown [6]. Our research seems to provide an answer to this question from a statistical point of view: CO, NO$_2$, O$_3$ combined with PM$_{10}$ and PM$_{2.5}$ from the day to a lag of 8 or 9 days have a significant correlation with COVID-19 occurrence; in contrast, only CO and NO$_2$ combined with PM$_{10}$ from the day to a lag of 9 days have a significant correlation with COVID-19 intensity, and except NO$_2$, other determinants that reflect air quality have exerted a positive impact. These differences may be due to the different trends of each determinant during the study period (Figure 3), and the determined relationship and the mechanism cause need more research. The unavailability of the atmospheric environmental quality data before January 1, 2020, which has led to the delay effect of air pollution after 9 days not being considered, but it may still work.

Conclusions

The COVID-19 pandemic’s diffusion patterns are definitively caused by a multiplicity of environmental, economic, social factors as well as public health control measures. Temperature has a limited impact on the spread of the epidemic, and the contain of COVID-19 requires scientific and systematic PHCMs. Our findings have deepened our understanding of the factors that contribute to the spread of the epidemic, and the experience of China could provide lessons for countries that are still experiencing a health emergency and help to prevent future pandemics similar to COVID-19.
Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Availability of data and material
The datasets used and/or analyzed during the current study are available from the corresponding author.

Competing interests
The authors declare that they have no competing interests.

Funding
This work was supported by the Hubei Provincial Natural Science Foundation of China (Grant No. 2020CFA048) and the National Natural Science Foundation of China (Grant No.: 62071457 and 81803297).

Authors' contributions
Juan Qiu and Rendong Li conceived and designed the study. Juan Qiu analyzed the data. Dongfeng Han, Qihui Shao, Yifei Han, Xiyue Luo and Yanlin Wu contributed to data collection and preprocessing. Juan Qiu and Rendong Li wrote the paper.

References
1. WHO Coronavirus (COVID-19) Dashboard [https://covid19.who.int/]
2. Altizer S, Ostfeld RS, Johnson PTJ, Kutz S, Harvell CD: Climate Change and Infectious Diseases: From Evidence to a Predictive Framework. Science 2013, 341(6145):514-519.
3. Thomas MB: Epidemics on the move: Climate change and infectious disease. Plos Biology 2020, 18(11).
4. Patz JA, Epstein PR, Burke TA, Balbus JM: Global climate change and emerging infectious diseases. Jama-Journal of the American Medical Association 1996, 275(3):217-223.
5. Shakil MH, Munim ZH, Tasnia M, Sarowar S: COVID-19 and the environment: A critical review and research agenda. Science of the Total Environment 2020, 745.
6. Espejo W, Celis JE, Chiang G, Bahamonde P: Environment and COVID-19: Pollutants, impacts, dissemination, management and recommendations for facing future epidemic threats. Science of the Total Environment 2020, 747.
7. Runkle JD, Sugg MM, Leeper RD, Rao YH, Matthews JL, Rennie JJ: Short-term effects of specific humidity and temperature on COVID-19 morbidity in select US cities. *Science of the Total Environment* 2020, 740.

8. Yao Y, Pan JH, Wang WD, Liu ZX, Kan HD, Qiu Y, Meng X, Wang WB: Association of particulate matter pollution and case fatality rate of COVID-19 in 49 Chinese cities. *Science of the Total Environment* 2020, 741.

9. Lian XB, Huang JP, Zhang L, Liu CW, Liu XY, Wang LN: Environmental Indicator for COVID-19 Non-Pharmaceutical Interventions. *Geophysical Research Letters* 2021, 48(2).

10. Copat C, Cristaldi A, Fiore M, Grasso A, Zuccarello P, Santo Signorelli S, Conti GO, Ferrante M: The role of air pollution (PM and NO2) in COVID-19 spread and lethality: A systematic review. *Environmental Research* 2020, 191.

11. Saez M, Tobias A, Barcelo MA: Effects of long-term exposure to air pollutants on the spatial spread of COVID-19 in Catalonia, Spain. *Environmental Research* 2020, 191.

12. Baker RE, Yang W, Vecchi GA, Metcalf CJE, Grenfell BT: Susceptible supply limits the role of climate in the early SARS-CoV-2 pandemic. *Science* 2020, 369(6501):315-319.

13. Accarino G, Lorenzetti S, Aloisio G: Assessing correlations between short-term exposure to atmospheric pollutants and COVID-19 spread in all Italian territorial areas. *Environmental Pollution* 2021, 268.

14. Kraemer MUG, Yang CH, Gutierrez B, Wu CH, Klein B, Pigott DM, du Plessis L, Faria NR, Li RR, Hanage WP *et al*: The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* 2020, 368(6490):493-+.

15. Jing QL, Liu MJ, Zhang ZB, Fang LQ, Yuan J, Mang AR, Dean NE, Luo L, Ma MM, Longini I *et al*: Household secondary attack rate of COVID-19 and associated determinants in Guangzhou, China: a retrospective cohort study. *Lancet Infect Dis* 2020, 20(10):1141-1150.

16. Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, Colaneri M: Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine* 2020, 26(6):855-+.

17. Coskun H, Yildirim N, Gunduz S: The spread of COVID-19 virus through population density and wind in Turkey cities. *Science of the Total Environment* 2021, 751.

18. Sarmadi M, Marufi N, Moghaddam VK: Association of COVID-19 global distribution and environmental and demographic factors: An updated three-month study. *Environmental Research* 2020, 188.

19. Gatto M, Bertuzzo E, Mari L, Miccoli S, Carraro L, Casagrandi R, Rinaldo A: Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. *Proceedings of the National Academy of Sciences of the United States of America* 2020, 117(19):10484-10491.

20. Huang JP, Zhang L, Liu XY, Wei Y, Liu CW, Lian XB, Huang ZW, Chou JF, Liu XR, Li X *et al*: Global prediction system for COVID-19 pandemic. *Science Bulletin* 2020, 65(22):1884-1887.

21. Zhao WW, Zhang JZ, Meadows ME, Liu YX, Hua T, Fu BJ: A systematic approach is needed to contain COVID-19 globally. *Science Bulletin* 2020, 65(11):876-878.

22. Zhao S: To avoid the noncausal association between environmental factor and COVID-19 when using aggregated data: Simulation-based counterexamples for demonstration. *Science of the Total Environment* 2020, 748.

23. Bontempi E, Vergalli S, Squazzoni F: Understanding COVID-19 diffusion requires an
interdisciplinary, multi-dimensional approach. *Environmental Research* 2020, 188.

24. Barouki R, Kogevinas M, Audouze K, Belesova K, Bergman A, Birnbaum L, Boekhold S, Denys S, Desseille C, Drakvik E *et al.*: The COVID-19 pandemic and global environmental change: Emerging research needs. *Environment International* 2021, 146.

25. Katoto P, Brand AS, Bakan B, Obadia PM, Kuhangana C, Kayembe-Kitenge T, Kitenge JP, Nkulu CBL, Vanoirbeek J, Nawrot TS *et al.*: Acute and chronic exposure to air pollution in relation with incidence, prevalence, severity and mortality of COVID-19: a rapid systematic review. *Environmental Health* 2021, 20(1).

26. Leung K, Wu JT, Liu D, Leung GM: First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment. *Lancet* 2020, 395(10233):1382-1393.

27. Tian HY, Liu YH, Li YD, Wu CH, Chen B, Kraemer MUG, Li BY, Cai J, Xu B, Yang QQ *et al.*: An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* 2020, 368(6491):638-+

28. Oraby T, Tyshenko MG, Maldonado JC, Vatcheva K, Elsaadany S, Alali WQ, Longenecker JC, Al-Zoughool M: Modeling the effect of lockdown timing as a COVID-19 control measure in countries with differing social contacts. *Scientific Reports* 2021, 11(1).

29. Highest public health alert for 30 provincial-level regions (Jan 25, 2020) [http://regional.chinadaily.com.cn/huzhou/2020-01/25/c_449939.htm]  

30. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) [https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6]

31. Xinjiang reports record spike in COVID-19 patients (2020-07-28) [https://www.chinadaily.com.cn/a/202007/28/W55f16f23ea31083481725c5ad.html]

32. Pang X, Ren L, Wu S, Ma W, Yang J, Di L, Li J, Xiao Y, Kang L, Du S *et al.*: Cold-chain food contamination as the possible origin of COVID-19 resurgence in Beijing. *National Science Review* 2020, 7(12):1861-1864.

33. Reasons for the Surge of Confirmed NCP Cases [https://www.chinadaily.com.cn/a/202002/13/WSS5e4613669a31012821727796e.html]

34. Wang J, Xie H, Fisher JF: *Multilevel Models: Applications using SAS*. Berlin: De Gruyter & Higher Education Press; 2011.

35. Tooze JA, Grunwald GK, Jones RH: Analysis of repeated measures data with clumping at zero. *Statistical Methods in Medical Research* 2002, 11(4):341-355.

36. Tibshirani R: Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society Series B-Methodological* 1996, 58(1):267-288.

37. Zhao P, Yu B: On model selection consistency of Lasso. *Journal of Machine Learning Research* 2006, 7:2541-2563.

38. Tibshirani R: The lasso method for variable selection in the cox model. *Statistics in Medicine* 1997, 16(4):385-395.

39. Alene M, Yismaw L, Assemie MA, Ketema DB, Gietaneh W, Birhan TY: Serial interval and incubation period of COVID-19: a systematic review and meta-analysis. *Bmc Infectious Diseases* 2021, 21(1).

40. Lauer SA, Grantz KH, Bi QF, Jones FK, Zheng QL, Meredith HR, Azman AS, Reich NG, Lessler J:
The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of Internal Medicine* 2020, 172(9):577-476.

Carteni A, Di Francesco L, Martino M: How mobility habits influenced the spread of the COVID-19 pandemic: Results from the Italian case study. *Science of the Total Environment* 2020, 741.

Wu XX, Yin J, Li CL, Xiang HX, Lv M, Guo ZY: Natural and human environment interactively drive spread pattern of COVID-19: A city-level modeling study in China. *Science of the Total Environment* 2021, 756.

Normalcy returns to Wuhan, but with masks, of course [https://global.chinadaily.com.cn/a/202008/26/WS5f45eb3da310675eafc55a0e.html]

Feng Zhiming, Tang Yan, Yang Yanzhao, Dan Z: The Relief Degree of Land Surface in China and Its Correlation with Population Distribution. *ACTA GEOGRAPHICA SINICA* 2007, 62(10):1073-1082.

Huanyong H: The distribution, regionalization and prospect of China’s population. *ACTA GEOGRAPHICA SINICA* 1990, 45(2):139-145.

Xia Haibin, Dai Xiaoye, Wang Ying, Zhen W: Analysis of the Relationship between Social Economic Development and Land Use in Hainan Province. *AREAL RESEARCH AND DEVELOPMENT* 2006, 25(3):120-124+130.

Liu L: Emerging study on the transmission of the Novel Coronavirus (COVID-19) from urban perspective: Evidence from China. *Cities* 2020, 103(102759).

Wang Jiaoe, Du Delin, Wei Ye, Haoran Y: The development of COVID-19 in China: Spatial diffusion and geographical pattern. *GEOGRAPHICAL RESEARCH* 2020, 39(7):1450-1462.

Peeri NC, Shrestha N, Rahman MS, Zaki R, Tan ZQ, Bibi S, Baghbanzadeh M, Aghamohammadi N, Zhang WY, Haque U: The SARS, MERS and novel coronavirus (COVID-19) epidemics, the newest and biggest global health threats: what lessons have we learned? *International Journal of Epidemiology* 2020, 49(3):717-726.

Duhon J, Bragazzi N, Kong JD: The impact of non-pharmaceutical interventions, demographic, social, and climatic factors on the initial growth rate of COVID-19: A cross-country study. *Science of the Total Environment* 2021, 760.

Cho SW: Quantifying the impact of nonpharmaceutical interventions during the COVID-19 outbreak: the case of Sweden. *Econometrica Journal* 2020, 23(3):323-344.

Li Y, Campbell H, Kulkarni D, Harpur A, Nundy M, Wang X, Nair H, Usher Network C-ER: The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2: a modelling study across 131 countries. *Lancet Infect Dis* 2021, 21(2):193-202.

Cacho PM, Hernandez JL, Lopez-Hoyos M, Martinez-Taboada VM: Can climatic factors explain the differences in COVID-19 incidence and severity across the Spanish regions?: An ecological study. *Environmental Health* 2020, 19(1).

Paraskevis D, Kostaki EG, Alygizakis N, Thomaidis NS, Cartalis C, Tsiodras S, Dimopoulos MA: A review of the impact of weather and climate variables to COVID-19: In the absence of public health measures high temperatures cannot probably mitigate outbreaks. *Science of the Total Environment* 2021, 768.

Ward MP, Xiao S, Zhang ZJ: The role of climate during the COVID-19 epidemic in New South
Wales, Australia. Transboundary and Emerging Diseases 2020, 67(6):2313-2317.

56. Sun YR, Hu XK, Xie J: Spatial inequalities of COVID-19 mortality rate in relation to socioeconomic and environmental factors across England. Science of the Total Environment 2021, 758.

57. Pirouz B, Haghshenas SS, Pirouz B, Haghshenas SS, Piro P: Development of an Assessment Method for Investigating the Impact of Climate and Urban Parameters in Confirmed Cases of COVID-19: A New Challenge in Sustainable Development. Int J Environ Res Public Health 2020, 17(8).

58. Juni P, Rothenbuhler M, Bobos P, Thorpe KE, da Costa BR, Fisman DN, Slutsky AS, Gesink D: Impact of climate and public health interventions on the COVID-19 pandemic: a prospective cohort study. Canadian Medical Association Journal 2020, 192(21):E566-E573.

59. Benjamin Bowe, Yan Xie, Andrew K. Gibson, Miao Cai, Aaron van Donkelaar, Randall V. Martin, Richard Burnett, Ziyad Al-Aly: Ambient fine particulate matter air pollution and the risk of hospitalization among COVID-19 positive individuals: Cohort study. Environment International 2021, 154(106564).

Fig. 1. The accumulated confirmed COVID-19 cases per 100,000 persons at the city level in mainland China from Jan 10 to Jun 11, 2020.

Fig. 2. Mean fixed (background) determinant distribution.

Fig. 3. Determinants that change over time.
Figure 1

The accumulated confirmed COVID-19 cases per 100,000 persons at the city level in mainland China from Jan 10 to Jun 11, 2020.
Figure 2

Mean fixed (background) determinant distribution.
Figure 3

Determinants that change over time.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementary.docx