



A multiplicity of environmental, economic and social factor analyses to understand COVID-19 diffusion

Juan Qiu^{a,*}, Rendong Li^a, Dongfeng Han^{a,b}, Qihui Shao^{a,b}, Yifei Han^{a,b}, Xiyue Luo^c, Yanlin Wu^c

^a 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, China

^b University of Chinese Academy of Sciences, Beijing, China

^c Faculty of Resources and Environmental Science, Hubei University, Wuhan, China

ARTICLE INFO

Keywords:

COVID-19
Meteorology
Atmospheric environmental quality
Human mobility
Non-pharmaceutical interventions
Mixed-effects mixed-distribution model

ABSTRACT

Research on the impact of the environment on COVID-19 diffusion lacks a full-comprehensive perspective, and neglecting the multiplicity of the human-environment system can lead to misleading conclusions. We attempted to reveal all pre-existing environmental-to-human and human-to-human determinants that influence the transmission of COVID-19. As such, 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 factors related to climate, atmospheric environmental quality, terrain, population, economic, human mobility as well as non-pharmaceutical interventions (NPIs). Not only the changes in determinants over time as the pandemic progresses but also their lag and interaction effects were examined. CO, O₃, PM₁₀ and PM_{2.5} were found 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. NPIs' 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. 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. The COVID-19 diffusion are caused by a multiplicity of environmental, economic, social factors as well as NPIs. First, multiple pollutants carried simultaneously on particulate matter affect COVID-19 transmission. Second, the temperature has a limited impact on the spread of the epidemic. Third, NPIs must last for at least 15 days or longer before the effect has been apparent. Fourth, the impact of population movement from the epicenter on COVID-19 gradually diminished over time and intraregional migration deserves more attention.

1. Introduction

The COVID-19 pandemic is still ongoing, as of August 2021. A better understanding of the effects of comprehensive natural (environmental-to-human and pollution-to-human) and human (human-to-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) [1–3], 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 refers to the factors accounting for pollution-to-human and environmental-to-human transmission mechanisms [4,5], including air pollutants such as PM_{2.5}, PM₁₀, SO₂, NO₂, O₃ and CO, and meteorological variables such as temperature and relative humidity [6–10]. In addition, the human environment refers to the factors accounting for human-to-human diffusion mechanisms [4,5]. What 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 [11–17]. Non-pharmaceutical

* Corresponding author.

E-mail address: qiujuan@apm.ac.cn (J. Qiu).

<https://doi.org/10.1016/j.onehlt.2021.100335>

Received 4 August 2021; Received in revised form 16 September 2021; Accepted 3 October 2021

Available online 5 October 2021

2352-7714/© 2021 The Authors.

Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

interventions (NPIs) 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 [11–13,16–18]. The pandemic is actually a very complex phenomenon, and its diffusion patterns are typically caused by a multiplicity of environmental, economic and social factors, hence, full-comprehensive variables and systematic analytical procedures should be considered in the analysis to describe any possible correlation in a rigorous way [10,19–22]. 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 [20,22].

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 [23]. 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 [24,25] that appear to be related to imported cases or seafood [26]. Therefore, the period from the first confirmed case in Wuhan to June 11, 2020, in China was a perfect real-world 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 conducted an evaluation in China on developing a comprehensive vision of COVID-19 contagion that considered the multiplicity of environmental-to-human and human-to-human channels. At the same time, the lag effects of NPIs, population

movement and atmospheric environment were also discussed.

2. Methods

2.1. 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. For the outlier processing, see Additional A and Fig. S1.

The association between an independent factor and disease incidence counts may fail to infer the association with disease transmission [19,22], 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.

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, $Day2 = Day * Day$.

Table 1 lists all the variables and their data sources.

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, NO₂,

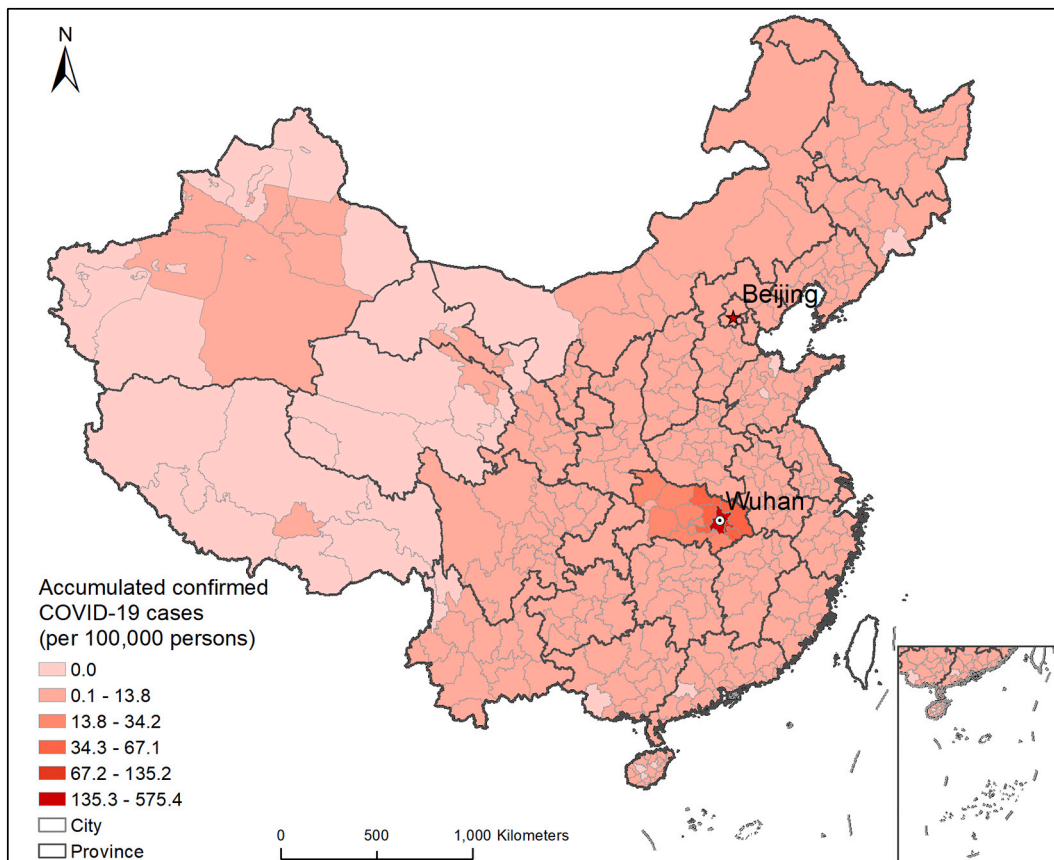


Fig. 1. The accumulated confirmed COVID-19 cases per 100,000 persons at the city level in mainland China from Jan 10 to Jun 11, 2020.

Table 1
Data type, variables and data source.

Data type	Variable	Abbreviation	Units	Scale	Data source		
Disease	COVID-19 data	Case incidence ratios (cases per 10 million persons)	CIR	0.00001%	Daily	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	
Natural environment	Climate data	Minimum temperature	MinT	0.1 °C	Daily	http://data.cma.cn	
		Maximum temperature	MaxT	0.1 °C	Daily		
		Mean temperature	MeanT	0.1 °C	Daily		
		Relative humidity	Rh	%	Daily		
		Minimum relative humidity	MinRh	%	Daily		
	Atmospheric environmental quality data	Carbon monoxide, CO	CO, CO_1, ..., CO_9	mg/m ³	Daily	https://datacenter.mee.gov.cn/	
		Nitrogen dioxide, NO ₂	NO2, NO2_1, ..., NO2_9	µg/m ³	Daily		
		Ozone, O ₃	O3, O3_1, ..., O3_9	µg/m ³	Daily		
		Fine particles, PM _{2.5}	PM25, PM25_1, ..., PM25_9	µg/m ³	Daily		
		Inhalable coarse particles, PM ₁₀	PM10, PM10_1, ..., PM10_9	µg/m ³	Daily		
Human environment	Terrain data	Sulfur dioxide, SO ₂	SO2, SO2_1, ..., SO2_9	µg/m ³	Daily	http://www.gscloud.cn	
		Air Quality Index	AQI, AQI_1, ..., AQI_9		Daily		
	Population data	Mean DEM	MeanDEM	m	Daily		http://tjj.shandong.gov.cn/tjnj/nj2020/zk/indexch.htm , etc.
		Household population	Pop	10,000 Person	2019, Fixed value		
	Economic data	Population density	PD	Person/km ²	2019, Fixed value		http://www.mohurd.gov.cn/xytj/index.html
		Gross domestic product per capita	GDP	100 million RMB	2019, Fixed value		
	Human mobility	Destination migration scale flow from Wuhan = destination proportion in population flow from Wuhan * migration scale	Popmob, Popmob1, ..., Popmob9, Popmobsum		Daily		http://qianxi.baidu.com/
		The distance of each city from Wuhan	DisWH	km	Fixed value		
	Public health control measures	National emergency response	Reslevel, Reslevel1, ..., Reslevel20		Daily		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

O₃, PM_{2.5}, PM₁₀, SO₂ 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 the 103 missing values of air pollutants and AQI. Besides, the delay effect of atmospheric environmental indicators was also identified. For example, CO₁ is defined as CO delayed by one day, etc., CO₉ is defined as CO delayed by 9 days. Other atmospheric environmental indicators are similar.

The terrain data ASTER GDEM 30 M 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, to test whether the effect of GDP on CIR changes over time as the pandemic progresses, the interaction effect of GDP and time was considered, it was named Day_GDP = Day * 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 Pobjob, Pobjob1 is defined as a delay of one day, etc., Pobjob9 is defined as a delay of 9 days. The cumulative Pobjob (Pobjobsum) 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 * 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 Additional file, Table S1. 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 ordinal categorical variables, specifically, 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.

2.2. Statistical analysis

The CIRs were extremely nonnormally distributed, exhibiting a large clump of values at zero (51776/56364, 91.8%) and skewed nonzero values (Additional Table S2, Figs. S2 and S3). To address these semi-continuous 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 [27,28]. 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 > 0), allowing for repeated measurements using random effects and allowing for correlation between the two components.

An SAS Macro, MIXCORR [28] 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):

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

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

where p_{ij} is the probability of having nonzero CIR values ($R_{ij} = 1$, R_{ij} represents the occurrence variable), CIR_{ij} is city i 's CIR on Day j , and X_{1kj} is a vector of covariates that explain $\text{Pr}(R_{ij} = 1)$. 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, Day² ($\text{Day}_{ij}^2 = \text{Day}_{ij} * \text{Day}_{ij}$), is constructed in the model. Regression coefficients β_{10} , β_{11} , β_{111} , and β_{1j} are the intercept, fixed time (Day) effect, fixed time (Day²) effect and fixed effects of covariates on the log-odds of $R_{ij} = 1$, respectively, and μ_{1j} is the random effect of individuals on the log-odds. In these multilevel modeling terms, ($\beta_{10} + \mu_{1j}$) is the random intercept that allows the probabilities of having a nonzero CIR to vary across cities.

CIR intensity model (a linear regression model):

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

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 μ_{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 μ_{1j} and μ_{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 [29], which has been proven to be useful and feasible when the number of observations is much larger than the number of predictors [30,31]. 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).

3. Results

3.1. 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 Pobjob0, Pobjob1, Pobjob2, Pobjob3, Pobjob4 and Pobjob5 were negative, and the slopes of variables Pobjob5, Pobjob5 and Pobjobsum were positive. Considering that the incubation period of COVID-19 ranges from 2 to 14 days or even longer [32,33], 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, Pobjobsum 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

Table 2

Lasso variable selection results. Estimate lasso with lambda = 1427.596 (lopt). Lasso is the variable coefficient estimated by Lasso, and Post-est is the variable coefficient of OLS regression on the selected variables.

Selected	Lasso	Post-est	Selected	Lasso	Post-est
Pop	-0.7565	-0.9184	SO2_7	-0.0283	-0.0421
PD	0.4994	0.5822	SO2_8	-0.0276	-0.0385
DisWH	-0.4102	-0.4808	SO2_9	-0.0551	-0.0557
GDP	0.9788	1.0951	NO2	-0.0762	-0.1032
MeanDEM	-0.0050	-0.0084	NO2_1	-0.0086	-0.0083
MinT	-0.0123	-0.0145	NO2_2	-0.0143	-0.0299
MinRh	-0.0120	-0.0330	NO2_3	-0.0046	-0.0082
Popmob0	-0.4786	-0.5358	NO2_4	-0.0027	-0.0061
Popmob1	-0.3684	-0.3436	NO2_5	-0.0087	-0.0125
Popmob2	-0.0497	-0.1001	NO2_6	-0.0231	-0.0287
Popmob3	-0.1051	-0.1209	NO2_8	-0.0149	-0.0223
Popmob4	-0.2876	-0.2543	NO2_9	-0.0059	0.0152
Popmob5	-0.0412	-0.2639	O3	-0.0145	-0.0204
Popmob7	0.0166	0.2176	O3_1	-0.0003	-0.0003
Popmob9	1.5534	1.5899	O3_2	-0.0043	-0.0042
Popmobsum	0.0152	0.0153	O3_6	-0.0049	-0.0057
Reslevel2	0.1354	0.0305	O3_7	-0.0006	-0.0006
Reslevel3	0.5769	0.4942	O3_9	-0.0028	-0.0044
Reslevel5	0.0473	-0.1455	PM10	0.0014	0.0022
Reslevel6	0.2925	0.3833	PM10_1	0.0006	0.0011
Reslevel8	0.5027	0.5793	PM10_2	0.0012	0.0014
Reslevel9	0.9803	1.5329	PM10_3	0.0007	0.0012
Reslevel15	-0.2266	-0.7235	PM10_4	0.0016	0.0020
Reslevel17	-0.1993	-0.2165	PM10_5	0.0006	0.0010
Reslevel20	-2.5200	-2.4571	PM10_6	0.0003	0.0007
CO	1.7358	2.9689	PM10_7	0.0011	0.0015
CO_2	0.0384	0.9589	PM10_9	0.0020	0.0038
CO_9	-0.1670	-2.0334	PM25_7	0.0009	0.0005
			PM25_8	0.0115	0.0149

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 (NO_{2_9}). The variables with the longest delay time were selected to enter the next modeling, such as CO₉, SO_{2_9}, NO_{2_9}, O_{3_9}, PM_{10_9} and PM_{25_8}. In

Table 3

Parameter estimates, standard errors and fit statistics for mixed-effects mixed-distribution model to CIR occurrence and intensity.

Parameter	Occurrence			Intensity		
	Estimate	SE	p	Estimate	SE	p
Intercept	-2.9447	0.2349	<0.0001	3.1101	0.1528	<0.0001
Day (β_{11}, β_{21})	0.1765	0.0068	<0.0001	-0.0115	0.0037	0.0018
Day2 (β_{111}, β_{211})	-0.0030	0.0001	<0.0001	-0.0003	0.0000	<0.0001
Pop	0.0618	0.0237	0.0096	-0.1784	0.0158	<0.0001
PD	0.0813	0.0244	0.0010	-0.0266	0.0167	0.1116
GDP	0.0460	0.0181	0.0114	0.0688	0.0113	<0.0001
Day_GDP	0.0023	0.0002	<0.0001	-0.0002	0.0001	0.0099
MeanDEM	-0.0052	0.0009	<0.0001	-0.0011	0.0007	0.0862
DisWH	-0.1865	0.0136	<0.0001	-0.0852	0.0096	<0.0001
Day_DisWH	0.0028	0.0002	<0.0001	0.0027	0.0002	<0.0001
Popmobsum	0.0019	0.0003	<0.0001	0.0026	0.0002	<0.0001
Reslevel15	-0.5810	0.0192	<0.0001	-0.1162	0.0095	<0.0001
MinT	-0.0008	0.0005	0.1106	-0.0004	0.0003	0.1979
MinRh	0.0026	0.0012	0.0350	-0.0031	0.0007	<0.0001
CO_9	0.3844	0.0820	<0.0001	0.0171	0.0516	0.7411
NO2_9	-0.0258	0.0029	<0.0001	-0.0101	0.0015	<0.0001
O3_9	0.0075	0.0010	<0.0001	0.0035	0.0006	<0.0001
PM10_9	0.0016	0.0004	0.0002	0.0015	0.0005	0.0021
PM25_8	0.0038	0.0007	<0.0001	0.0005	0.0004	0.1564
SO2_9	0.0013	0.0039	0.7423	-0.0002	0.0026	0.9424
Residual(σ_e^2)				0.5681	0.0123	<0.0001
Random effect(σ_1^2, σ_2^2)	0.9592	0.1021	<0.0001	0.4287	0.0416	<0.0001
Covariance ($\rho_{\sigma_1\sigma_2}$)	0.2819	0.0529	<0.0001			
Fit statistics	Uncorrelated	Correlated	Difference in -2 log likelihood			p
AIC	44,686.5	44,634.17				
-2 Log Likelihood	44,600.5	44,546.17	54.33			<0.0001

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.

3.2. 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 NPIs 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 NPIs had a statistically significant effect on the probability of CIR occurrence, except MinT ($p = 0.1106$) and SO_{2_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 CO₉, O_{3_9}, PM_{10_9} and PM_{25_8} and negatively associated with MeanDEM, DisWH, Reslevel15 and NO_{2_9}. Furthermore, the interaction of GDP, DisWH and Day (Day_GDP and Day_DisWH, respectively) was positively correlated with CIR occurrence. However, eleven variables and interaction effects had a statistically significant effect on CIR intensity, except PD, MeanDEM, MinRh and PM_{25_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, PM_{2.5} 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 ($\hat{\sigma}_1^2 = 0.9592, p < 0.0001; \hat{\sigma}_2^2 = 0.4287, p < 0.0001$)

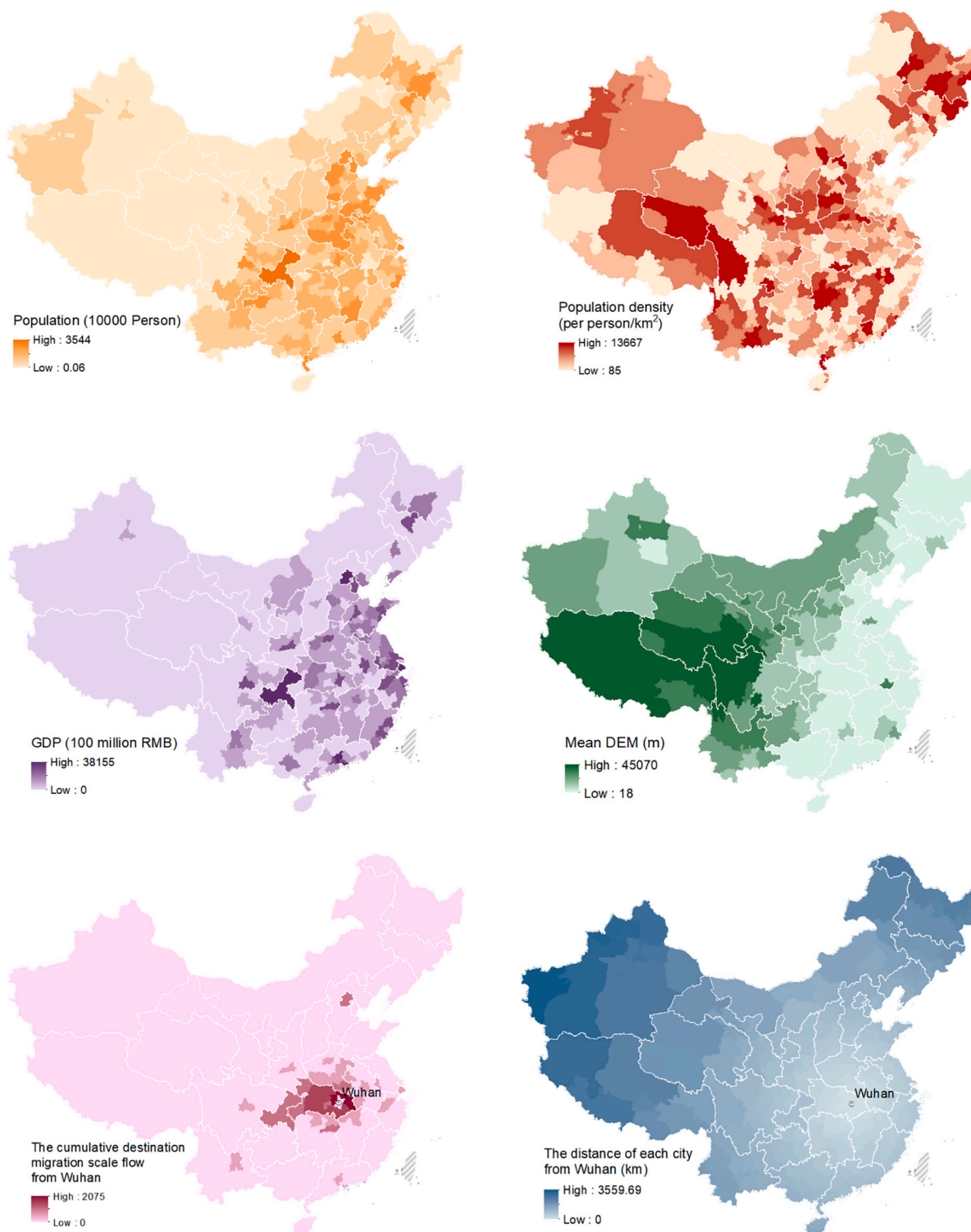


Fig. 2. Mean fixed (background) determinant distribution.

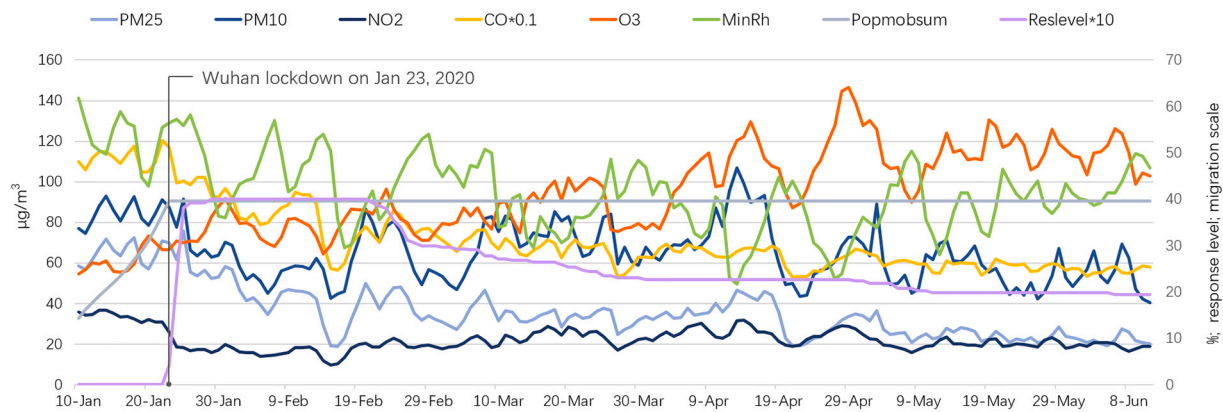


Fig. 3. Determinants that change over time.

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.

4. 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.

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 [34] but is not associated with accumulated COVID-19 cases [35]. The difference in those results was due mainly to the different choice of dependent variables reflecting the COVID-19 epidemic index [22]. 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 larger CIR occurrence and intensity (Table 3). Prior research found that the number of COVID-19 cases is higher in high-income countries [14]. 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 [36] 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 [37–39].

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). Being farther away from Wuhan was a protective factor of COVID-19 transmission. 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.

The positive correlation between CIR and the level of NPIs 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 [40], 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 [41], and the effect of introducing on reproduction number was delayed by 1–3 weeks at the country level [42].

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 [43], while there have been studies that support the absence of any correlation, low sensitivity [44] or even a positive one [45]. 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 NPIs.

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 [46] and mortality rate [47], and a positive correlation or negative correlation was found between relative humidity and confirmed cases in different study areas in Italy [48], while the association between absolute humidity and epidemic growth was no longer significant [49]. 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 [50]; 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 [7]. Our research seems to provide an answer to this question from a statistical point of view: CO, NO₂, O₃ combined with PM₁₀ 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 NO₂ combined with PM₁₀ from the day to a lag of 9 days have a significant correlation with COVID-19 intensity, and except NO₂, other determinants that reflect air quality have exerted a positive impact. Although the relationship between NO₂ and COVID-19 epidemiological indicators is inconsistent, it is generally believed that NO₂ and other air pollutants increase the risk of COVID-19, especially in terms of fatality [7,8]. Our results on the relationship between air pollutants and CIR may be due to the different trends of each determinant during the study period (Fig. 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.

There are still some limitations in terms of data sources. First, we did not consider social acceptance of restriction measures since it is generally high in China. Second, in our study period, the epidemic was mainly affected by the domestic spreading while imported cases were excluded, and thus GDP was remained to be an influencing factor but the international trade and commercial exchange were ignored.

5. Conclusions

The COVID-19 pandemic's diffusion patterns are definitively caused by a multiplicity of environmental, economic, social factors and NPIs. First, multiple pollutants carried simultaneously on particulate matter act in a synergistic way to affect COVID-19 transmission. Second, the effect of temperature on COVID-19 was confounded by other determinants. Third, the impact of population movement from the epicenter on COVID-19 gradually diminished over time and intraregional migration deserves more attention. Fourth, NPIs must last for at least 15 days before the effect has been apparent. These findings could provide lessons for countries that are still experiencing a health emergency and help to prevent future pandemics similar to COVID-19.

Authors' contributions

JQ and RL conceived and designed the study. JQ analyzed the data. DH, QS, YH, XL and YW contributed to data collection and preprocessing. JQ and RL wrote the paper.

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).

Data availability

The dataset we generated has been uploaded the GitHub project (<https://github.com/JuanQiu2021/COVID-19Envdata>).

Declaration of Competing Interest

We declare no competing interests.

Acknowledgements

Not applicable.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.onehlt.2021.100335>.

References

- [1] S. Altizer, R.S. Ostfeld, P.T.J. Johnson, S. Kutz, C.D. Harvell, Climate change and infectious diseases: from evidence to a predictive framework, *Science* 341 (6145) (2013) 514–519, <https://doi.org/10.1126/science.1239401>.
- [2] M.B. Thomas, Epidemics on the move: climate change and infectious disease, *PLoS Biol.* 18 (11) (2020), <https://doi.org/10.1371/journal.pbio.3001013>.
- [3] M.G. Walsh, S. Sawleshwarkar, S. Hossain, S.M. Mor, Whence the next pandemic? The intersecting global geography of the animal-human interface, poor health systems and air transit centrality reveals conduits for high-impact spillover, *One Health* 11 (2020), <https://doi.org/10.1016/j.onehlt.2020.100177>.
- [4] U. Anand, C. Cabrerros, J. Mal, F. Ballesteros, M. Sillanpaa, V. Tripathi, E. Bontempi, Novel coronavirus disease 2019 (COVID-19) pandemic: from transmission to control with an interdisciplinary vision, *Environ. Res.* 197 (2021), <https://doi.org/10.1016/j.envres.2021.111126>.
- [5] E. Bontempi, Commercial exchanges instead of air pollution as possible origin of COVID-19 initial diffusion phase in Italy: more efforts are necessary to address interdisciplinary research, *Environ. Res.* 188 (2020), <https://doi.org/10.1016/j.envres.2020.109775>.
- [6] M.H. Shakil, Z.H. Munim, M. Tasnia, S. Sarowar, COVID-19 and the environment: a critical review and research agenda, *Sci. Total Environ.* 745 (2020), <https://doi.org/10.1016/j.scitotenv.2020.141022>.
- [7] W. Espejo, J.E. Celis, G. Chiang, P. Bahamonde, Environment and COVID-19: pollutants, impacts, dissemination, management and recommendations for facing future epidemic threats, *Sci. Total Environ.* 747 (2020), <https://doi.org/10.1016/j.scitotenv.2020.141314>.
- [8] C. Copat, A. Cristaldi, M. Fiore, A. Grasso, P. Zuccarello, S. Santo Signorelli, G. O. Conti, M. Ferrante, The role of air pollution (PM and NO₂) in COVID-19 spread and lethality: a systematic review, *Environ. Res.* 191 (2020), <https://doi.org/10.1016/j.envres.2020.110129>.
- [9] R.E. Baker, W. Yang, G.A. Vecchi, C.J.E. Metcalf, B.T. Grenfell, Susceptible supply limits the role of climate in the early SARS-CoV-2 pandemic, *Science* 369 (6501) (2020) 315–319, <https://doi.org/10.1126/science.abc2535>.
- [10] G. Accarino, S. Lorenzetti, G. Aloisio, Assessing correlations between short-term exposure to atmospheric pollutants and COVID-19 spread in all Italian territorial areas, *Environ. Pollut.* 268 (2021), <https://doi.org/10.1016/j.envpol.2020.115714>.
- [11] M.U.G. Kraemer, C.H. Yang, B. Gutierrez, C.H. Wu, B. Klein, D.M. Pigott, L. du Plessis, N.R. Faria, R.R. Li, W.P. Hanage, et al., The effect of human mobility and control measures on the COVID-19 epidemic in China, *Science* 368 (6490) (2020) 493, <https://doi.org/10.1126/science.abb4218>.
- [12] Q.L. Jing, M.J. Liu, Z.B. Zhang, L.Q. Fang, J. Yuan, A.R. Mang, N.E. Dean, L. Luo, M.M. Ma, I. Longini, et al., Household secondary attack rate of COVID-19 and associated determinants in Guangzhou, China: a retrospective cohort study, *Lancet Infect. Dis.* 20 (10) (2020) 1141–1150, [https://doi.org/10.1016/s1473-3099\(20\)30471-0](https://doi.org/10.1016/s1473-3099(20)30471-0).
- [13] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, M. Colaneri, Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy, *Nat. Med.* 26 (6) (2020) 855, <https://doi.org/10.1038/s41591-020-0883-7>.
- [14] M. Sarmadi, N. Marufi, V.K. Moghaddam, Association of COVID-19 global distribution and environmental and demographic factors: an updated three-month study, *Environ. Res.* 188 (2020), <https://doi.org/10.1016/j.envres.2020.109748>.
- [15] M. Gatto, E. Bertuzzo, L. Mari, S. Miccoli, L. Carraro, R. Casagrandi, A. Rinaldo, Spread and dynamics of the COVID-19 epidemic in Italy: effects of emergency containment measures, *Proc. Natl. Acad. Sci. U. S. A.* 117 (19) (2020) 10484–10491, <https://doi.org/10.1073/pnas.2004978117>.
- [16] L. Huang, X. Zhang, A. Xu, Effectiveness of interventions as part of the One Health approach to control coronavirus disease 2019 and stratified case features in Anhui Province, China: a real-world population-based cohort study, *One Health* 12 (2021), <https://doi.org/10.1016/j.onehlt.2021.100224>.
- [17] E.K. Yeoh, K.C. Chong, C.J. Chiew, V.J. Lee, C.W. Ng, H. Hashimoto, S. Kwon, W. Wang, N.N.S. Chau, C.H.K. Yam, et al., Assessing the impact of non-pharmaceutical interventions on the transmissibility and severity of COVID-19 during the first five months in the Western Pacific Region, *One Health* 12 (2021), <https://doi.org/10.1016/j.onehlt.2021.100213>.
- [18] H.Y. Tian, Y.H. Liu, Y.D. Li, C.H. Wu, B. Chen, M.U.G. Kraemer, B.Y. Li, J. Cai, B. Xu, Q.Q. Yang, et al., An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China, *Science* 368 (6491) (2020) 638, <https://doi.org/10.1126/science.abb6105>.
- [19] S. Zhao, To avoid the noncausal association between environmental factor and COVID-19 when using aggregated data: simulation-based counterexamples for demonstration, *Sci. Total Environ.* 748 (2020), <https://doi.org/10.1016/j.scitotenv.2020.141590>.

- [20] E. Bontempi, S. Vergalli, F. Squazzoni, Understanding COVID-19 diffusion requires an interdisciplinary, multi-dimensional approach, *Environ. Res.* 188 (2020), <https://doi.org/10.1016/j.envres.2020.109814>.
- [21] R. Barouki, M. Kogevinas, K. Audouze, K. Belesova, A. Bergman, L. Birnbaum, S. Boekhold, S. Denys, C. Desseille, E. Drakvik, et al., The COVID-19 pandemic and global environmental change: emerging research needs, *Environ. Int.* 146 (2021), <https://doi.org/10.1016/j.envint.2020.106272>.
- [22] G.H. Kerr, H.S. Badr, L.M. Gardner, J. Perez-Saez, B.F. Zaitchik, Associations between meteorology and COVID-19 in early studies: inconsistencies, uncertainties, and recommendations, *One Health* 12 (2021), <https://doi.org/10.1016/j.onehlt.2021.100225>.
- [23] Highest Public Health Alert for 30 Provincial-level Regions. http://regional.chinadaily.com.cn/huzhou/2020-01/25/c_449939.htm, Jan 25, 2020.
- [24] 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>].
- [25] Xinjiang Reports Record Spike in COVID-19 Patients. <https://www.chinadaily.com.cn/a/202007/28/WS5f1f623ea31083481725c5ad.html>, 2020-07-28.
- [26] X. Pang, L. Ren, S. Wu, W. Ma, J. Yang, L. Di, J. Li, Y. Xiao, L. Kang, S. Du, et al., Cold-chain food contamination as the possible origin of COVID-19 resurgence in Beijing, *Natl. Sci. Rev.* 7 (12) (2020) 1861–1864, <https://doi.org/10.1093/nsr/nwaa264>.
- [27] J. Wang, H. Xie, J.F. Fisher, *Multilevel Models: Applications Using SAS*, De Gruyter & Higher Education Press, Berlin, 2011.
- [28] J.A. Tooze, G.K. Grunwald, R.H. Jones, Analysis of repeated measures data with clumping at zero, *Stat. Methods Med. Res.* 11 (4) (2002) 341–355, <https://doi.org/10.1191/09622802sm291ra>.
- [29] R. Tibshirani, Regression shrinkage and selection via the Lasso, *J. R. Stat. Soc. Ser. B* 58 (1) (1996) 267–288, <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>.
- [30] P. Zhao, B. Yu, On model selection consistency of Lasso, *J. Mach. Learn. Res.* 7 (2006) 2541–2563.
- [31] R. Tibshirani, The lasso method for variable selection in the cox model, *Stat. Med.* 16 (4) (1997) 385–395, [https://doi.org/10.1002/\(sici\)1097-0258\(19970228\)16:4<385::aid-sim380>3.0.co;2-3](https://doi.org/10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3).
- [32] M. Alene, L. Yismaw, M.A. Assemie, D.B. Ketema, W. Gietaneh, T.Y. Birhan, Serial interval and incubation period of COVID-19: a systematic review and meta-analysis, *BMC Infect. Dis.* 21 (1) (2021), <https://doi.org/10.1186/s12879-021-05950-x>.
- [33] S.A. Lauer, K.H. Grantz, Q.F. Bi, F.K. Jones, Q.L. Zheng, H.R. Meredith, A.S. Azman, N.G. Reich, J. Lessler, The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application, *Ann. Int. Med.* 172 (9) (2020) 577, <https://doi.org/10.7326/m20-0504>.
- [34] H. Coskun, N. Yildirim, S. Gunduz, The spread of COVID-19 virus through population density and wind in Turkey cities, *Sci. Total Environ.* 751 (2021), <https://doi.org/10.1016/j.scitotenv.2020.141663>.
- [35] X.X. Wu, J. Yin, C.L. Li, H.X. Xiang, M. Lv, Z.Y. Guo, Natural and human environment interactively drive spread pattern of COVID-19: a city-level modeling study in China, *Sci. Total Environ.* 756 (2021), <https://doi.org/10.1016/j.scitotenv.2020.143343>.
- [36] Normalcy returns to Wuhan, but with masks, of course [<https://global.chinadaily.com.cn/a/202008/26/WS5f45eb3da310675eaf55a0e.html>].
- [37] Feng Zhiming, Tang Yan, Yanzhao Yang, Z. Dan, The relief degree of land surface in China and its correlation with population distribution, *Acta Geograph. Sin.* 62 (10) (2007) 1073–1082, <https://doi.org/10.3321/j.issn:0375-5444.2007.10.007>.
- [38] H. Huanyong, The distribution, regionalization and prospect of China's population, *Acta Geograph. Sin.* 45 (2) (1990) 139–145, <https://doi.org/10.3321/j.issn:0375-5444.1990.02.003>.
- [39] Xia Haibin, Dai Xiaoye, Wang Ying, W. Zhen, Analysis of the relationship between social economic development and land use in Hainan province, *Areal Res. Dev.* 25 (3) (2006) 120–124, 130, <https://doi.org/10.3969/j.issn.1003-2363.2006.03.027>.
- [40] J. Duhon, N. Bragazzi, J.D. Kong, The impact of non-pharmaceutical interventions, demographic, social, and climatic factors on the initial growth rate of COVID-19: a cross-country study, *Sci. Total Environ.* 760 (2021), <https://doi.org/10.1016/j.scitotenv.2020.144325>.
- [41] S.W. Cho, Quantifying the impact of nonpharmaceutical interventions during the COVID-19 outbreak: the case of Sweden, *Econ. J.* 23 (3) (2020) 323–344, <https://doi.org/10.1093/ectj/utaa025>.
- [42] Y. Li, H. Campbell, D. Kulkarni, A. Harpur, M. Nundy, X. Wang, H. Nair, C.-E. R. Usher Network, 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.* 21 (2) (2021) 193–202, [https://doi.org/10.1016/s1473-3099\(20\)30785-4](https://doi.org/10.1016/s1473-3099(20)30785-4).
- [43] P.M. Cacho, J.L. Hernandez, M. Lopez-Hoyos, V.M. Martinez-Taboada, Can climatic factors explain the differences in COVID-19 incidence and severity across the Spanish regions?: an ecological study, *Environ. Health* 19 (1) (2020), <https://doi.org/10.1186/s12940-020-00660-4>.
- [44] M. Su, S.S. Peng, L.L. Chen, B. Wang, Y. Wang, X.R. Fan, Z.M. Dong, A warm summer is unlikely to stop transmission of COVID-19 naturally, *Geohealth* 4 (12) (2020), <https://doi.org/10.1029/2020gh000292>.
- [45] D. Paraskevis, E.G. Kostaki, N. Alygizakis, N.S. Thomaidis, C. Cartalis, S. Tsiodras, M.A. Dimopoulos, 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, *Sci. Total Environ.* 768 (2021), <https://doi.org/10.1016/j.scitotenv.2020.144578>.
- [46] M.P. Ward, S. Xiao, Z.J. Zhang, The role of climate during the COVID-19 epidemic in New South Wales, Australia, *Transbound. Emerg. Dis.* 67 (6) (2020) 2313–2317, <https://doi.org/10.1111/tbed.13631>.
- [47] Y.R. Sun, X.K. Hu, J. Xie, Spatial inequalities of COVID-19 mortality rate in relation to socioeconomic and environmental factors across England, *Sci. Total Environ.* 758 (2021), <https://doi.org/10.1016/j.scitotenv.2020.143595>.
- [48] B. Pirouz, S.S. Haghshenas, B. Pirouz, S.S. Haghshenas, P. Piro, 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* 17 (8) (2020), <https://doi.org/10.3390/ijerph17082801>.
- [49] P. Juni, M. Rothenbuhler, P. Bobos, K.E. Thorpe, B.R. da Costa, D.N. Fisman, A. S. Slutsky, D. Gesink, Impact of climate and public health interventions on the COVID-19 pandemic: a prospective cohort study, *Can. Med. Assoc. J.* 192 (21) (2020) E566–E573, <https://doi.org/10.1503/cmaj.200920>.
- [50] 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, *Environ. Int.* 154 (106564) (2021), <https://doi.org/10.1016/j.envint.2021.106564>.